Gene-level consequences of new cancer-specific chromosomal rearrangements

Kaja Beate Nyquist

Oslo, September 2013

Thesis for the degree of Philosophiae Doctor (PhD)

UiO : University of Oslo

Institute for Clinical Medicine
Faculty of Medicine
University of Oslo

Oslo University Hospital

Section for Cancer Cytogenetics
Institute for Medical Informatics
The Norwegian Radium Hospital

Centre for Cancer Biomedicine

Centre for Cancer Biomedicine
Norwegian Centre of Excellence
University of Oslo

NORWEGIAN CANCER SOCIETY

The Norwegian Cancer Society

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1733

ISBN 978-82-8264-717-5

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

Gene-level consequences of new cancer-specific chromosomal rearrangements

Kaja Beate Nyquist

University of Oslo

2013

Cover illustration

My first love in science was the stars and planets, and for some years I dreamt about becoming an astronaut and exploring the universe above. In cancer research I found a starry sky in the FISH microscope, where bright signals in different colors resembled the stars I had dreamt about as a child. Even though we now know more about the components inside the body, sometimes it feels as if we have a whole universe within where only few stars have been fully characterized.

Table of Contents

Acknowledgements

The work presented in this thesis was carried out at the Section for Cancer Cytogenetics, Institute for Medical Informatics at the Norwegian Radium Hospital in Oslo. The studies were financed by grants from the South-Eastern Regional Health Authority and the Norwegian Cancer Society.

First of all, I would like express my gratitude to my supervisor Francesca Micci for her guidance and support through these years. You have always being enthusiastic and encouraging me in my work, even when we met hurdles along the way. I am grateful for the thoughtfulness you have had for me, both professionally and privately.

I am also very grateful to my co-supervisor professor Sverre Heim for introducing me to the world of cancer cytogenetics. Your immerse knowledge in the field has been a true inspiration, and you have opened a word of fuzzy black and white creatures I didn't know existed. Thank you for your patience and constructive advice in the writing process, your contribution has been invaluable.

A special thanks to Jim Thorsen. I am grateful for your patience trying to teach me anything from how to optimize the PCR cycle to make figures with vector-based graphics. Your time and effort has been truly appreciated. Ioannis Panagopoulos who joined us in 2012, also deserves a note. Your impressive knowledge and experience in the field of fusion genes have been inspirational to me, thank you for sharing your knowledge.

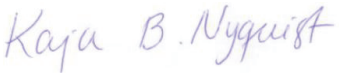
I am grateful to all of my co-authors for each of their individual contributions. I want to mention in particular Bodil Bjerkehagen and Anne Tierens at the Pathology Department. Thank you for always being helpful and for taking the time to advise me in a busy clinical setting.

To all the girls at the lab; thank you for a pleasant working atmosphere and interesting discussions during lunch. A special thanks to Lisbeth Haugom and Anne Mette Eng Eibak. Thank you for teaching me the lab procedures, and for always being helpful.

Thank you to my former colleagues in the Ph.D forum- for expanding my view of the PhD experience and cancer research in general.

A warm thanks to my colleagues at the “stipendiat-office” Thank you for sharing frustrations and achievements, for inspiring conversations, for “English Wednesday” and “Bun-Friday”. It has been a pleasure to share office with you, and I know I will miss you.

Finally I want to thank my family and friends for taking interest in my work. To Karsten and Edla for being the best family I could dream of. A special though to my deceased mother in law Gabriella who put my view on cancer research perspective, and who I know would have loved to see this work finalized.



Oslo, September 2013

List of papers

Paper I

Kaja B. Nyquist, Jim Thorsen, Bernward Zeller, Alf Haaland, Gunhild Trøen, Sverre Heim, Francesca Micci:

Identification of the *TAF15-ZNF384* fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12).

Cancer Genetics (2011), 204(3):147-152.

Paper II

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Roberta Roberto, Hilde S. Wik, Anne Tierens, Sverre Heim, Francesca Micci:

t(12;13)(q14;q31) leading to *HMGA2* upregulation in acute myeloid leukaemia.

British Journal of Haematology (2012), 157(6):769–771.

Paper III

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Lisbeth Haugom, Ludmila Gorunova, Bodil Bjerkehagen, Alexander Fosså, Marianne Guriby, Torfinn Nome, Ragnhild A. Lothe, Rolf I. Skotheim, Sverre Heim, Francesca Micci.

Whole-transcriptome sequencing identifies novel *IRF2BP2-CDX1* fusion gene brought about by translocation t(1;5)(q42;q32) in mesenchymal chondrosarcoma.

PLOS ONE (2012), 7(11):e49705.

Paper IV

Francesca Micci, Jim Thorsen, Ioannis Panagopoulos, **Kaja B. Nyquist**, Bernward Zeller, Anne Tierens, Sverre Heim:

High-throughput sequencing identifies an *NFIA/CBFA2T3* fusion gene in acute erythroid leukemia with t(1;16)(p31;q24).

Leukemia (2013) 27(4):980–982.

Introduction

Cancer and cytogenetics–from mysterious lumps to chromosomal barcodes

One of the first known descriptions of cancer is found in a papyrus which scientists have dated to approximately 2500 BC (Mukherjee, 2011). It is believed to be the teachings of the Egyptian physician Imhotep and describes patients with various medical conditions. One description tells about a woman with a hard, bulging mass of the breast, without fever or discharging fluids. This is most likely a description of a woman with breast cancer. In addition to written reports, solid evidence of cancer has been detected in mummified bodies with bone tumors (Mukherjee, 2011). These ancient findings imply the coexistence of man and neoplasia for as long as we have been here.

Throughout history, likewise, many theories of the origin of cancer have been proposed. Hippokrates (460-370 BC), referred to by many as the father of Western medicine, adhered to the dominant theory of his time believing that the body had four fluids or humors (black bile, yellow bile, phlegm, and blood) and that cancer was caused by an imbalance affecting predominantly the black bile (American Cancer Society, 2009). This humoral theory more or less held sway until the mid nineteenth century when the German pathologist Rudolph Virchow established cellular pathology as the new paradigm of disease, and especially cancerous diseases. One of Virchow's assistants, David Hansemann, studied mitoses in several human tumor types and documented mitotic asymmetries in most of them; this he realized might point to a mechanism of carcinogenesis. This line of thinking matured into the somatic mutation theory of cancer suggested by Theodor Boveri in 1914. His theory was based on studies of fertilized sea-urchin eggs, originally enough, but led to our current standard understanding of the principles of tumorigenesis (Balmain, 2001). Boveri observed that eggs fertilized with two sperm cells sometimes developed an unequal distribution of chromosomes in the daughter cells. Some of these cells died, but some lived on with an abnormal appearance. Boveri therefore suggested that individual chromosomes must possess different abilities, and that some abnormal combinations of

chromosomes could create abnormal growth. In retrospect, his assumptions fit surprisingly well with the theory of proto-oncogenes and tumor suppressor genes discovered over 70 years later (Balmain, 2001). It is today generally accepted that cancer is a genetic disease caused by DNA sequence changes (Stratton *et al*, 2009).

The next major contribution to cancer genetics came in 1960 when Nowell and Hungerford identified a minute chromosome, called the Philadelphia chromosome, in the bone marrow cells of patients suffering from chronic myeloid leukemia (CML) (Nowell and Hungerford, 1960b). This was the first specific chromosome abnormality to be detected in cancer. Cytogenetic analysis experienced a boom in the 1970s after the development of banding protocols for human chromosomes (Caspersson *et al*, 1968). Until then chromosomes had been arranged according to their size in 23 matching pairs and grouped after the location of their centromere (Trask, 2002). The new staining techniques revealed dark and light bands across the arms of the chromosomes, and these soon became the bands of reference for classifying normal and aberrant chromosomes. Modern cancer cytogenetics is an important tool for an initial screening of the cancer genome, both in diagnostics and research. Cytogenetic studies have greatly improved our understanding of carcinogenesis of especially hematological and mesenchymal neoplasias (Heim and Mitelman, 2009).

The *BCR-ABL* story

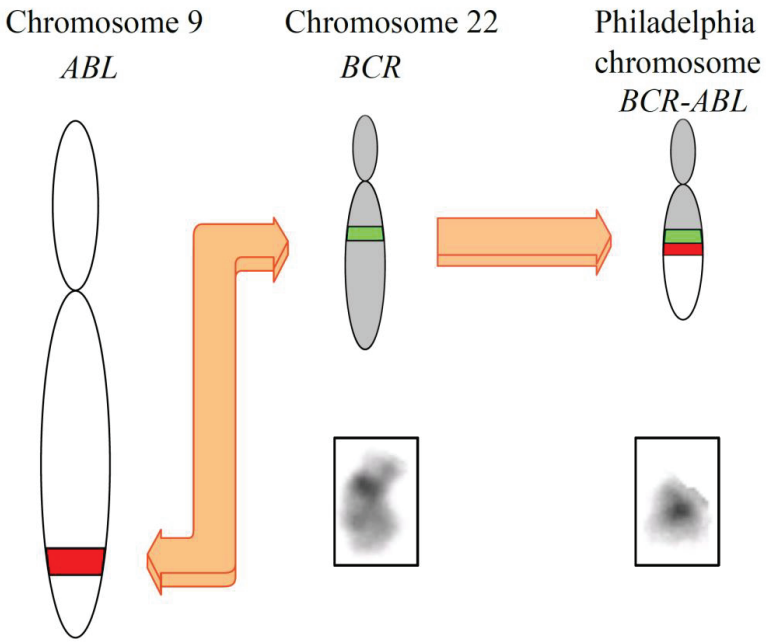


Figure 1: The formation of the *BCR-ABL* fusion gene is caused by a translocation of the long arm of chromosome 9 to chromosome 22. Below the schematic presentations is a photo of the normal chromosome 22 and the smaller "Philadelphia" chromosome.

Nowell and Hunderfords's finding, the so called "Philadelphia chromosome" in patients with CML (Nowell and Hungerford, 1960b) (Figure 1), turned out to be one of the most seminal in the history of cancer genetics. This chromosome, named after the city in which it was discovered, is specific for chronic myeloid leukemia. In the early 1970s, it was discovered that it resulted from a translocation between chromosomes 9 and 22 (Rowley, 1973, Goldman, 2010). On chromosome 9, the *ABL1* gene was involved (de Klein *et al*, 1982). In 1984, the breakpoint on chromosome 22 was shown to occur within a 5.8 Mb area called the breakpoint cluster region (today the *BCR* gene) (Groffen *et al*, 1984). The fusion of these two genes results in a qualitatively new gene that encodes an aberrant tyrosine kinase which is crucial for CML development (Lugo *et al*, 1990, Deininger *et al*, 2000). Based on this knowledge,

Druker and colleagues searched for compounds that could specifically bind and inhibit the *BCR-ABL1* protein (Druker *et al*, 1996). Their finding, a tyrosine kinase inhibitor (TKI) called imatinib, has since revolutionized treatment of CML patients (Druker, 2002). Although imatinib does not cure the patients of their disease, it gives long-term remissions without significant side effects in the majority of patients, and it is today the first choice of treatment for newly diagnosed patients with chronic phase CML (Baccarani *et al*, 2006).

Targeted cancer therapy

Targeted therapy aims to interfere with a molecular target essential for cancer cell growth and survival. In contrast to classical chemotherapy, targeted therapy aims to kill cancer cells more selectively and thus avoid the adverse side effects and general toxicity of cytostatics (Sawyers, 2004). A potential target for therapy should be (1) necessary and sufficient for cancer cell survival, (2) expressed and active in the tumor, (3) inhibiting effector molecules for which no possible circumventions in the implied pathway is possible (Borden *et al*, 2003). Inhibition with imatinib of the mutant kinase chimeric protein BCR-ABL1 in CML is the most successful example of this approach, even though resistance ultimately will develop (see above). However, in most other cancers where TKIs have been tested, patients develop resistance within a shorter time frame (Sierra *et al*, 2010). Efforts are now being made to try to understand the mechanisms underlying resistance both *in vitro* and *in vivo*.

Cytogenetic nomenclature

Tjio and Levan detected in the mid 1950s (Tjio and Levan, 1956) the correct human chromosome number of 46 (44 autosomes and 2 sex chromosomes) after it had been believed for many years that 48 was the correct number. In the years after, research in this field expanded and a common system for cytogenetic nomenclature was needed. The first generally accepted cytogenetic nomenclature was established at a meeting in Denver in 1960, and this became the forerunner of today's International System for Cytogenetic Nomenclature (ISCN) (Shaffer *et al*, 2009). Cytogenetic nomenclature is based upon the chromosomal banding pattern and the occurrence of other specific, distinct morphological hallmarks. Chromosomes are arranged by their size and the location of the centromeres. Accordingly, chromosome 1 is the largest chromosome, chromosome 2 the second largest and so on, with one exception, that chromosome 21 is smaller than chromosome 22 (Shaffer *et al*, 2009). The chromosomal bands appear by the differential staining throughout the chromosomes and are defined as "a part of a chromosome that is clearly distinguishable from its adjacent segments by appearing darker or lighter with one or more banding techniques"(Shaffer *et al*, 2009). The shorter chromosome arm is designated the "p" arm, while the longer arm is called the "q" arm. The chromosomal location 2p13 therefore designates chromosome 2, the short arm, region 1 within that arm, and band 3 within that region (Figure 2).

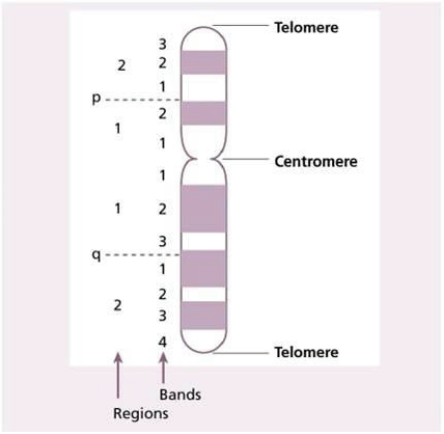


Figure 2: Graphic presentation of a chromosome. See text for details. From: Essential Haematology, 6th Edn. © Hoffbrand A.V. and Moss P.A.H., Blackwell Publishing, 2011 (with permission)

Nomenclature of abnormal chromosomes

Chromosomal aberrations are divided in two groups after their nature: structural and numerical. Structural aberrations are rearrangements of one or more chromosomes, whereas numerical aberrations refer to copy number changes (Heim and Mitelman, 2009). The most common structural aberrations are translocations, deletions, and inversions. They can be balanced, with no net gain or loss of genetic material, or unbalanced, implying unequal exchange of genetic material. Several other abnormalities can be seen and a specific terminology is used for karyotypic descriptions. A list of cytogenetic abbreviations can be found in Table 1. Only clonal chromosomal abnormalities are reported in karyotypic descriptions. A clone is defined as a group of cells stemming from a single progenitor cell, and at least two cells with the same structural aberration must be detected in order to infer the existence of a clone when doing cytogenetic analyses (Shaffer *et al*, 2009). If the aberration is numerical, that is loss or gain of a whole chromosome, two cells are needed for a supernumerary chromosome to be considered clonal, and three cells if the chromosome is lost. Clonally related cells share one or more aberrations, but additional abnormalities can be detected as a part of clonal evolution.

Rearrangement	Abbreviation	Description
Addition	add	Addition of material of unknown chromosomal origin
Composite	cp	Composite karyotype describes all clonally occurring abnormalities in a cytogenetically heterogeneous tumor
Deletion	del	Deletion (terminal or interstitial); loss of chromosomal material seemingly including or not including the tip of the chromosome arm
Derivative	der	Structurally rearranged chromosome
Marker	mar	Abnormal chromosome that cannot be identified
Translocation	t	Interchanges of material between two non-homologous chromosomes

Table 1: Cytogenetic abbreviations relevant for this thesis

Consequences of balanced chromosomal translocations

Balanced chromosomal translocations involve swapping of material between two chromosomes with no net gain or loss of chromosomal material. Two main pathogenetic consequences may in the cancer context follow from such translocations: either two genes are fused with each other and a qualitatively new hybrid or fusion gene is formed, or one gene loses or gains regulatory elements leading to its aberrant expression (Mitelman *et al*, 2007) (Figure 3). Translocations can also be present without any known consequences. Some of these are probably non-recurrent changes seen in unstable cancer genomes, but others might involve as yet undiscovered mechanisms leading to growth advantages for the cancer cell.

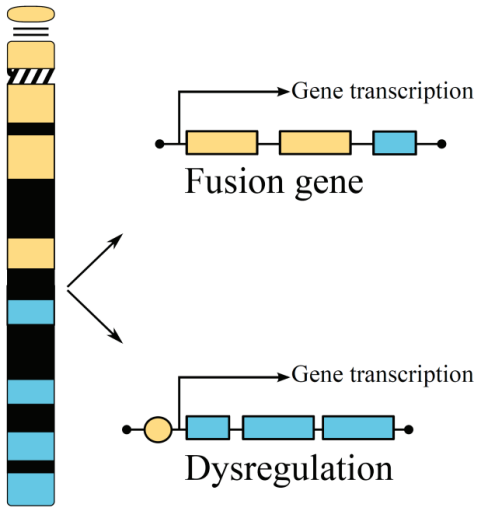


Figure 3: The two main consequences of chromosomal translocations are illustrated; the formation of a fusion gene and the dysregulation of a normal gene. The bars indicate exons while the circle illustrates a regulatory element.

Fusion genes

Fusion genes are hybrid genes formed by the fusing together of two previously separate genes. If joined properly, the new gene may give rise to a new chimeric protein with novel and oncogenic properties different from those of the native proteins. Based on cytogenetic studies, it has been estimated that fusion genes account for about 20% of cancer morbidity (Mitelman *et al*, 2007), but this may well be only the tip of the iceberg. Most known fusion genes are found in leukemias, lymphomas and mesenchymal tumors, while fewer have so far been found in the more common epithelial cancers. This does not necessarily mean that fusion genes in fact are more common in these rare tumor types, it may rather reflect the number of cytogenetically analyzed cases or the ease with which fusion genes are found in different settings (Mitelman *et al*, 2004). Studies relying on new investigative technologies are now revealing fusion genes on a much larger scale than was previously acknowledged,

especially in epithelial cancers (Edwards, 2010). The number of known fusion genes in cancer has increased from about 100 in the year 2000 (Mitelman, 2000) to 1603 as of August 14, 2013 (Mitelman *et al*, 2013). Recurrent fusion genes have now been detected also in common epithelial cancers such as carcinoma of the prostate (Tomlins *et al*, 2005), lung (Soda *et al*, 2007), breast (Stephens *et al*, 2009), and colon (Bass *et al*, 2011). Some of the cases analyzed revealed multiple fusion genes within the same tumor, with hitherto unknown oncogenic potential (Maher *et al*, 2009). In mesenchymal tumors, a fusion gene is often found in a high proportion of cases within each diagnostic entity. By comparison, the fusion genes identified so far in carcinomas have been found in a much lower percentage of cases.

Several aspects make the fusion genes of cancer interesting to study and some of them will be discussed below. For many of the well-described fusion genes there is convincing evidence for an important role in cancer cell survival and initiation (Mitelman *et al*, 2007). Identification of new fusion genes may therefore locate new players in the complex process of carcinogenesis. In addition, the detection in a tumor sample of a specific fusion gene can help resolve diagnostic dilemmas. An example is the finding of *FUS-DDIT3* in a lipomatous tumor. This means that the tumor is a myxoid liposarcoma resulting from a t(12;16)(q13:p11); over 80% of myxoid liposarcomas harbor this translocation (Fletcher *et al*, 2013). This is not a unique example; in fact, most fusions are specific for certain cancer types. Certain gene fusions occur across several cancer types. *ETV6-NTRK1* is such a fusion gene, which has been detected in secretory breast cancer (Tognon *et al*, 2002), congenital fibrosarcoma (Knezevich *et al*, 1998), and acute myeloid leukemia (Eguchi *et al*, 1999). Phenotypic tumor classifications will therefore always be needed. Furthermore, and maybe most interestingly, the fusion genes may represent therapeutic targets unique to the tumor in question. The main problem with current chemotherapeutics is their lack of specificity. The drugs kill all dividing cells, causing severe side-effects to the patient and reducing the dosage the patients can tolerate. Again, the imatinib story demonstrates the success of this approach. Imatinib treatment has increased survival and is generally well tolerated by CML patients (Druker, 2008).

Gene deregulation

If a chromosomal rearrangement does not lead to the formation of a fusion gene in a cancer, chances are good that it juxtaposes regulatory sequences from one gene with the transcribing parts of the other gene, which usually is a proto-oncogene, and therefore gives rise to the latter gene's activation. As the DNA structure is three-dimensional, also promoters far from the translocation breakpoint can affect gene transcription. One well-known example of gene deregulation following a translocation is the juxtaposition of the proto-oncogene *MYC*, from 8q24, to the vicinity of an immunoglobulin gene. This deregulation is typical of Burkitt lymphoma (Hecht and Aster, 2000) and leads to constitutive overexpression of *MYC*. In other cases, genes at translocation breakpoints may be inactivated by the chromosomal rearrangement. This seems to be rare, and the functional consequences are often difficult to predict (Mitelman *et al*, 2007), however the result could be loss of tumor suppressor function. The affected genes can either be translated into truncated transcripts or be degraded as a consequence of the cellular quality control system for example by a mechanism called nonsense-mediated decay that recognizes preterm stop-codons (Chang *et al*, 2007). The result would be an aberrant protein with unknown functional consequences, in the former case, and lack of the protein in the latter.

The human miRNA genes, which are small, non-coding RNA molecules with influence on transcriptional and post-transcriptional regulation of gene expression, are non-randomly located at hot spots for chromosomal rearrangements (Calin *et al*, 2004). Translocations in these areas may lead to deregulation of protein coding genes in a manner similar to that seen in lymphomas (Calin and Croce, 2007). Also, genes can contain binding sites for miRNAs with repressive functions. Disruption of these sites through chromosomal translocations may lead to overexpression of the affected genes. An example of this mechanism is found in translocations involving 12q13-15 affecting the *HMGA2* gene (Mayr *et al*, 2007) (paper II).

Rearrangements without functional consequences

Rearrangements may also occur in non-coding areas of DNA with no functional gene alteration resulting. The rarity of reports describing this phenomenon may be due to

publication bias, i.e., journals are less likely to publish negative results. The possibility is well illustrated by individuals carrying a balanced translocation with apparently normal phenotypes. Approximately 1 of 1000 newborns are estimated to carry such a translocation (Ferlin *et al*, 2007), and because all transcribed genes are intact, no functional consequence is apparent. The translocation might first be detected when an affected individual turns out to be infertile or have spontaneous abortions because of a higher likelihood of genomically unbalanced offspring.

Submicroscopic rearrangements around breakpoints

Not all seemingly balanced translocations turn out to be exactly balanced when analyzed at the base pair level. Deletions (Sinclair *et al*, 2000), inversions (Desmaze *et al*, 1997, Micci *et al*, 2006), paper III), and duplications (Howarth *et al*, 2011) in and around the translocation breakpoints may accompany what looks cytogenetically as balanced changes, both in malignant diseases and in constitutional translocations. This possibility should be kept in mind when searching for fusion genes; it could account for otherwise difficult-to-explain findings made during the study. In addition, the true prevalence of this phenomenon might be underestimated as such a rearrangement makes the characterization of the breakpoint more difficult.

Acute leukemias and sarcomas

Hematological and mesenchymal malignancies were the targets of studies leading up to this thesis. Both originate from the mesoderm and in both types of neoplasia, balanced translocations leading to fusion genes are often identified (Mitelman *et al*, 2007).

The term leukemia (from Greek leucos: white, and haima: blood) was first used by the German pathologist Rudolf Virchow in 1847 (Kampen, 2012). He believed the disease to be caused by too many white blood cells. Today, leukemia is known to be a clonal neoplastic disease of hematopoietic cells. By way of operational definition, acute leukemia is said to be present when the bone marrow contains >20 percent immature cells or blasts. Depending on the leukemic cell lineage commitment, the acute leukemias are divided into acute lymphoblastic leukemia (ALL) if the cells resemble

immature lymphatic cells (in which case they are assumed to originate from cells of lymphoid origin) and, correspondingly, acute myeloid leukemia (AML) if made up of immature myeloid cells (Swerdlow, 2008). The term “acute” in contrast to “chronic” originally referred to the rapid course of the acute form of the disease; in current times, it reflects the immature nature of the cells of the acute leukemias. Acute leukemia can be fatal within days to weeks if not adequately treated. Signs of leukemia are caused by the replacement of the normal bone marrow cells by a malignant leukemic clone, and include manifestations brought about by anemia, thrombocytopenia, and neutropenia. As a consequence, patients with leukemia usually present with fatigue, bone pain, headache, infections, bleedings or other unspecific signs and symptoms.

Leukemias have in later decades been classified according to the FAB (French-American-British) system first launched in 1976 (Bennett *et al*, 1976) which is based on morphological and cytochemical leukemic cell features. Since 2001, the World Health Organization (WHO) classification has largely superseded previous classifications (Swerdlow, 2001). In addition to cell morphology, the WHO system includes immunophenotype and genetic characteristics in the classification.

ALL is divided in two main subgroups, those expressing B cell and those expressing T cell features, whereas AML is divided in six subgroups depending on the maturation level and the predominantly affected cell type. While ALL is primarily a disease of children with most cases being diagnosed in patients under 18 years of age (Swerdlow, 2008), AML is much more common in adults with a median age at diagnosis of about 70 years (Smith *et al*, 2011).

Known risk factors for the development of leukemias, especially the myeloid ones, include exposure to organic solvents such as benzene (Galbraith *et al*, 2010) and previous treatment with chemotherapeutic drugs (Czader and Orazi, 2009), but in the majority of cases no etiologic factor can be identified (Zeeb and Blettner, 1998).

When treating patients with acute leukemia, the first goal is to obtain a complete remission using chemotoxic agents. Whether this can be accomplished depends on how well the patient tolerates the induction treatment, their performance status at

diagnosis, co-morbidities, and biological properties of the neoplastic disorder. Cytogenetic abnormalities, along with other criteria, determine the risk classification and therefore indirectly influence the further treatment in each group (Vardiman *et al*, 2009). In the medium and high risk groups, hematopoietic cell transplantation during first remission is an option for young patients with good performance status, but the treatment is dependent of the existence of a HLA-matched donor and associated with several adverse side effects.

Sarcomas (from Greek sarx: flesh, and oma: tumor) are tumors replicating connective tissue architectural features and which presumably originate from cells of mesenchymal origin. They are anatomically subdivided into bone and soft tissue tumors (Fletcher *et al*, 2013). Sarcomas occur in all age-groups and usually present as a hard, painless swellings. They account for only 1 % of human malignant neoplasms (Fletcher *et al*, 2013). Sarcomas can be further subdivided into over 70 different histological subtypes according to the WHO (Fletcher *et al*, 2013). Another way of classifying sarcomas, in principle independently of their histological appearance and tissue of origin, is based on their cytogenetic changes (Taylor *et al*, 2011). Tumors in the first group harbor simple chromosomal rearrangements or near-diploid karyotypes while those of the other group typically display complex karyotypes with both numerical and structural rearrangements. In sarcomas with simple karyotypes (about a third of all sarcomas), fusion genes are often detected (Mitelman, 2000). Some of these fusion genes are tightly linked to a histological disease entity, as for example Ewing sarcoma. In this malignancy, the majority of cases display a t(11;22) leading to a fusion between the *EWSR1* and the *FLI1* genes (Delattre *et al*, 1992, Fletcher *et al*, 2013). Some sarcomas harbor instead activating point mutations in key genes, with gastrointestinal stromal tumors (GISTs) as the most thoroughly studied example. The majority of GISTs have a mutated tyrosine kinase gene which renders them susceptible to TKI treatment (Hirota *et al*, 1998, Heinrich *et al*, 2003).

Most sarcomas arise without any known etiology, but some genetic syndromes such as the Li-Fraumeni syndrome (Li *et al*, 1988) and neurofibromatosis type I (Ferner and

Gutmann, 2002) predispose to sarcoma development, as does previous radiation therapy (Bjerkehagen *et al*, 2008).

Limb-sparing surgery can be sufficient for cure in some patients with small tumors and no metastasis at diagnosis, but for most patients, multimodal approaches with combinations of chemotherapy and radiotherapy together with surgery are offered (Fletcher *et al*, 2013). Due to the rarity of sarcomas and especially tumors of individual subgroups, even multicenter trials spend years to include enough patients to conduct randomized trials for evaluation of new treatments. Therefore, it is difficult to obtain well-founded protocols. In research settings, it is possible to try new drugs based on a theoretical rationale for example in terminal ill patients. The first GIST patient treated with imatinib (Joensuu *et al*, 2001) demonstrated the potential success of this approach. A shift from a classification based on histological appearance to a genetic signature in these rare cancers may facilitate the introduction of new drugs.

Methods

The techniques described here are the main methods used in the studies behind this thesis. All methods have advantages as well as limitations, and some of them will be discussed in the following, together with a brief overview of the theoretical background underlying each method.

Fluorescence *in situ* hybridization

Fluorescence in situ hybridization (FISH) is a procedure based on the principle that complimentary DNA or RNA sequences hybridize to each other. The first hybridization using radioactively labeled DNA probes was described in the late 1960s (Pardue and Gall, 1969). Further development of hybridization techniques came in the early 1980s, after the discovery of the usefulness of fluorescent tags for labeling. These dyes are safer and simpler to handle, furthermore they can be stored for longer periods (Trask, 2002). These advantages led to several new applications for the method: characterization of chromosomal rearrangements, detection of microdeletions, and prenatal diagnosis of common congenital aneuploidies (Fan, 2002).

Three main types of FISH probes exist: locus-specific, whole chromosome painting, and repetitive sequence probes. Locus-specific probes hybridize to a defined area of interest in the genome. Whole chromosome painting probes contain a pool of DNA fragments all hybridizing on the same chromosome so that it appears fluorescent and can be easily identified in a metaphase. The repetitive sequence probes map to an amplified sequence in the genome. An example of the latter is the centromeric probe targeting pericentromeric sequences. They are well suited for detection of numerical aberrations either in interphase nuclei or if metaphase quality is poor (Moyzis *et al*, 1987).

The resolution of the experimental FISH approaches rapidly improved. By 1985 the first human gene, thyroglobulin, was mapped to a chromosome using cosmid clones of about 20 kb (Landegent *et al*, 1985). Improvements in software, hardware, probe technology, and reduced costs had a great impact on implementing the method, and today probes for FISH can detect areas down to 1 kb (Speicher and Carter, 2005).

One of the fascinating technological developments within the family of FISH methods is the so-called multicolor FISH. It is a whole genome screening technique using human chromosome painting probes to distinguish individual chromosomes. The spectral signature of each probe identifies the 24 different human chromosomes. Diverse imaging systems are available as, by way of example, M-FISH (Speicher *et al*, 1996) and Spectral karyotyping (Schrock *et al*, 1996). The method enables a relatively simple way of analyzing complex karyotypic rearrangements (Lee *et al*, 2001, Trask, 2002).

Another application of the FISH techniques came in 1992 through studies made by the group of Kallioniemi (Kallioniemi *et al*, 1992). The method is called comparative genomic hybridization (CGH) and is based on a competitive hybridization of tumor and normal DNA, labeled with different colors (green and red, respectively). The labeled DNA is then hybridized to normal metaphases. By measuring the green-to-red ratio it is possible to identify imbalances present in the tumor genome. Regions gained/amplified in cancer will appear green, whereas regions lost will appear red. The resolution of this chromosomal-based approach is about 10 Mb. Even higher resolution is obtained using array-based CGH. The first arrays had a resolution of about 40 kb (Pinkel *et al*, 1998). In this variant of competitive hybridization, tumor and normal DNA targets well-characterized probes attached to a glass slide. Later technological improvements made it possible to reduce the probe size and increase the probe number so that today all known exon sequences can be evaluated in one single hybridization experiment (Kapur *et al*, 2007). The progress of FISH methods elegantly shows the transition from classical cytogenetics to molecular analyses.

FISH with locus-specific probes derived from bacterial artificial chromosomes

DNA constructs of various sizes can be transfected to bacteria that function as stable vectors. Bacterial artificial chromosomes (BACs) are such constructs of unique sequences of about 150-300 kb in size (Shizuya *et al*, 1992). The BAC clones commercially available are either end-sequenced or fully sequenced. Clones mapping to the region of interest can easily be identified using different databases such as the UCSC Genome Browser (available at http://genome.ucsc.edu/). The BACs used in this

thesis were retrieved from the RPCI-11 library (Osoegawa *et al*, 2001) and the CalTech human BAC library (https://bacpac.chori.org/). As several genes can map within one BAC clone, the discrimination among them can be performed by FISH using smaller fosmid clones (Kim *et al*, 1992). After culturing of BAC and/or fosmid clones on an agar plate including antibiotics (the host bacteria have an antibiotic-resistant insert), the plasmids can be extracted by lysis of the bacterial cell membrane and separation of proteins and bacterial DNA. To label the extracted DNA, we used nick translation (Rigby *et al*, 1977), a technique whereby a fraction of the dNTPs in the DNA target strand is substituted with fluorescently marked nucleotides using DNAse and DNA polymerase 1. Denaturation of the slide and probe can be done either by heat or using chemicals such as formamide. After denaturation of both probe and target DNA on the slide, the probe will anneal to its target sequence because of its presence in higher numbers. After incubation in a humid chamber, a series of washes remove unspecific hybridization and slides are counterstained, typically using 4',6-diamidino-2-phenylindole (DAPI). The result of the hybridization is evaluated using a fluorescent microscope (Figure 4).

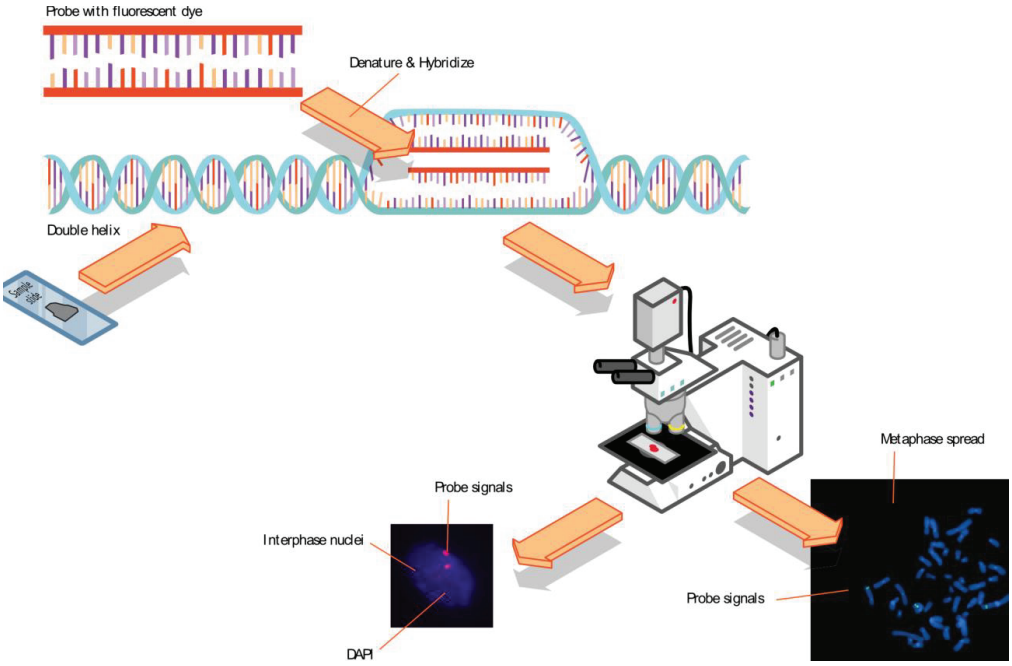


Figure 4: Overview of the FISH workflow

Metaphase FISH

The location of locus-specific FISH probes can be determined looking at only one metaphase. In a normal diploid cell, two signals will appear using a locus-specific probe. If a chromosomal break occurs in the sequence covered by the probe, three signals will be detected, one from the normal chromosome and one from each of the derivative chromosomes corresponding to a splitting of the second signal. In this manner, candidate genes involved in translocations can be detected by FISH. However, if the break occurs close to the end of the probe sequence, the extra signal might be too small to be detected. Metaphase FISH requires cell culturing and is dependent of the growth of malignant cells *in vitro*. If an aberration is recognized by karyotyping and material is sparse, it is possible to destain the very same slide and utilize it for FISH (Teixeira *et al*, 2000, Micci *et al*, 2002). Repeated hybridizations to the same slide testing different BAC probes are also possible (Epstein *et al*, 1995).

Interphase FISH

FISH experiments performed on interphase cells do not require cell culturing which may at times be a major advantage. FISH can then be performed on paraffin embedded material (Werner *et al*, 1997). Interphase FISH (IP-FISH) is often used for the detection of numerical aberrations as well as amplifications of suspected targets (Speicher and Carter, 2005). Commercially available kits exist for leukemias and solid tumors that can detect the most common aberrations. Because many cells are examined, this method can detect low numbers of genetically distinct subclones (Raimondi, 2000, Teixeira, 2002). But since the technique may be prone to false positive and/or false negative results, it is recommended to count a large number of nuclei and to test the specificity of the probe on a normal slide. In our laboratory, signals from 200 nuclei are counted by two independent observers and an abnormal signal pattern present in >10 % of cells is considered significant (Brandal *et al*, 2004). The probe design can reduce the number of false positive signals, for example by designing a double fusion probe. IP-FISH is, however, not suitable for breakpoint detection because of its inability to determine the precise probe localization.

Some limitations in the use of FISH on cancer cell preparations must be pointed out. Unspecific hybridization to both intrachromosomal and extrachromosomal areas may in some instances resemble specific hybridization and confuse the researcher. This limitation must be kept in mind especially when interpreting IP-FISH results. To overcome this problem, the procedure should be optimized and control tests performed in order to get reliable results. Prior to hybridization experiments, the target slides must be treated to remove proteins and other components that could interfere with probe hybridization.

Molecular methods for identifying and characterizing breakpoint regions

The minimal size of a FISH probe in regular fluorescence microscopy is about 40 kb. This resolution level is sufficient to identify genes involved in a rearrangement, but does not shed light on which parts of the genes are involved. To analyze in detail the breakpoint position, one has to resort to molecular methods. All methods described below are, to some extent, based on the polymerase chain reaction (PCR), first

described in a clinical setting by Saiki (Saiki *et al*, 1985) after its development by Kary Mullis (Bartlett and Stirling, 2003). The wide implementation of this method has revolutionized analyses in molecular biology. The PCR reaction is a method for amplifying double stranded DNA using sequential heating and cooling steps. Key elements in the reaction are primers, DNA polymerase, and nucleotides. Selective and repeated amplification is enabled by the sequence specific primers and the tolerance of the polymerase enzyme of high temperatures. The procedure continues in an exponential fashion using the newly formed DNA as template. The two used PCR primers are oriented in opposite directions and typically amplify a fragment of about 500-600 bp. In cases with a low number of target templates, a modified version of the PCR reaction called *nested* PCR can be applied. In this protocol, the sensitivity of the reaction is increased by running a second PCR with primers placed within the ones previously used. The procedure can be helpful in detecting rare mutations in a heterogeneous cell population. PCR based methods are extremely sensitive, therefore great care must be taken to avoid DNA contamination. PCR reactions searching for rare fusion genes are, however, less likely to produce false positives because of the rarity of the fusion genes we look for. Nonspecific binding of the primers to DNA or RNA can yield unspecific bands on PCR. To avoid this problem, careful primer design is important. The primers should among other factors be unique in the genome, should not tend to form secondary structures and the primer set should have a similar melting point (Sambrook and Russel, 2001).

Long-distance PCR

Under standard PCR reaction conditions, fragments up to two kb are amplified (Sambrook and Russel, 2001). Using modifications of the PCR protocol, genomic areas of up to 35 kb can be amplified (Barnes, 1994). To enable the synthesis of such long products, optimization of extension time, primer design, buffers, template quality, and the performance of the polymerase is necessary to obtain a good result. Given two candidate genes likely to take part in a translocation, long-distance PCR (LD-PCR) can be a good procedure to detect fusion genes, run with genomic DNA as the template. Primers are designed for each gene/exon and, if successful, the fusion gene is

amplified. If no bands occur, no information is obtained about the location of the rearrangement. Therefore, this method requires a clear idea as to where the breakpoints are situated.

Rapid amplification of cDNA ends

Rapid amplification of cDNA ends (RACE-PCR) can be applied if RNA is available and one of the fusion partners is identified. This method enables amplification of the 3`end of an mRNA transcript, and can thereby identify the second, previously unknown fusion partner. Using this method, only 23-25 base pairs of known sequence are necessary to run a PCR amplifying also the 3`end of the potential fusion partner. The method takes advantage of the construction of human mRNAs. These consist of the coding sequence followed by a poly-A tail (Scotto-Lavino *et al*, 2006a). Briefly, cDNA sequence is generated by reverse transcription of mRNA. The primer used targets the poly A tail and is universal for all mRNAs. In a second round of PCR, a gene–specific primer mapping to the gene of interest is used to amplify the target sequence. A nested primer located further downstream can give more specific products in a new PCR. The obtained PCR product is then sequenced and alignment with the reference genome identifies the unknown sequence. 5`end RACE is somewhat more complicated and requires modifications of the protocol because the Poly A tail cannot be used as template (Sambrook and Russell, 2001, Scotto-Lavino *et al*, 2006b). As the poly A tail is often destroyed during RNA degradation, RACE PCR requires RNA of good quality.

Quantitative PCR

In order to establish the expression level of an mRNA transcript, quantitative PCR (qPCR) is today the most frequently used and generally preferred method among researchers/molecular laboratories. The procedure is composed of three steps: (1) the conversion of RNA to cDNA using reverse transcriptase, (2) amplification of the cDNA template in a PCR reaction, and (3) detection of amplified products in real-time (Gibson *et al*, 1996). The method differs from conventional PCR by monitoring the amplification process whereas in conventional PCR only the end result is analyzed. There are currently two ways to follow the amplification process, both of which

measure increases in fluorescence. The TaqMan method (Livak *et al*, 1995) utilizes specifically labeled primer probes that, when binding to their target sequence, emit a fluorescence signal. The increase in target sequences as the PCR progresses results in intensified emission. The other systems use a more unselective probe that binds all double stranded DNA (for example, SYBR Green®), and a rise in DNA results in increased emission. Unselective amplification, for example from primer-dimer formation, is more problematic with this approach.

Possible pitfalls to qPCR are manifold and must be avoided from the very beginning, even before the sample is loaded on the qPCR instrument. Incorrect handling of the tumor tissue after it is removed from the human body can greatly impact the expression levels of the mRNAs. The RNA should be of good quality and, if more samples are compared, should be of similar quality. Unequal quantities of tumor cells in each sample may affect the expression levels of cancer specific genes. The method for RNA quality measurement should be equal for all samples as estimates of quality can differ from method to method. Furthermore, care must be taken when designing primers for an experiment. Traditionally, primers overlapping on one or two exons have been used as an expression level for the whole gene (Bustin *et al*, 2009). This can be problematic as most human genes can be alternatively spliced (de la Grange *et al*, 2007). Also, using sequence from only one exon may detect contaminating DNA. It is therefore recommended to use primers overlapping at least 2 exons (Bustin *et al*, 2009). Choosing controls for normalization is a challenging task. For a reference gene to be suitable for normalization, it should be expressed at a similar level across all samples analyzed, and must not be affected by the experimental conditions. The last step in the qPCR is data analysis, which is also very important. Spurious differences can occur if researches use different algorithms to interpret the data. Two main methods are used: absolute and relative quantification.

Sequencing

Since the 1970s, Sanger sequencing (Sanger *et al*, 1977) has been used to determine the order of bases in a given DNA fragment. The method consists of three main steps. First, the DNA fragment must be amplified, either by cloning into a high-copy-number

plasmid, or with primers in a PCR reaction. Then, cycles of template denaturation, primer annealing, and primer extension are carried out. This process resembles the PCR reaction apart for one important step: in addition to normal nucleotides, fluorescently labeled dideoxynucleotides are added to the reaction. Incorporation of these analogues results in a premature stop of the amplification process, and the resulting reaction mix will contain fragments of various lengths. By electrophoretic separation of the single-stranded end products, the sequence can be determined based on its emission spectrum. Sanger sequencing is applicable with DNA sequences up to about 1000 bases.

Since 2005, "next generation sequencing" has emerged and provides a faster and cheaper sequencing technology than the Sanger method. Advances in technology include improvements of the initial cloning step and simultaneous detection of numerous DNA fragments (Shendure and Ji, 2008). The technology is still evolving, and promising methods based on microchips reduce costs further. One of the main challenges using these new technologies is processing of the enormous amount of data generated (Shendure and Ji, 2008). This issue does not only apply to analysis of the data and storage capacity, but one also has to protect patient confidentiality in a safe way. In addition, the newer technologies yield shorter sequence reads, with lower accuracy than Sanger sequencing. Several applications for next-generation sequencing have emerged. Planning a given study, the approach chosen should reflect its aim. Sequencing platforms exist for genome-wide DNA studies, RNA studies, and targeted areas of the genome (Shendure and Lieberman Aiden, 2012). "High throughput", "long-read" and "bench top" instruments allow the researcher to choose between different costs and quality of the data obtained.

Transcriptome sequencing

The transcriptome is the complete set of transcripts of a cell, and RNA sequencing (RNA-seq) is a technique that analyzes the transcriptome through deep-sequencing (Wang *et al*, 2009). The first step in RNA sequencing after isolation of total RNA is construction of a cDNA library using mRNA as a template. mRNA must be isolated from the other RNA components (such as rRNA, tRNA siRNA, and miRNA).

Following purification, mRNA is fragmented and converted to single stranded cDNA. After second strand synthesis, the fragments are ligated using specific adapters enabling PCR amplification to a cDNA library. This library can then be sequenced using next generation sequencing.

After high quality reads have been obtained from sequencing of the cDNA, the reads must be aligned to a reference sequence for analysis and interpretation. Different algorithms for analysis have been developed for various purposes. We have used two algorithms for fusion gene detection in the studies described in this thesis, deFuse (McPherson *et al*, 2011) and FusionMap (Ge *et al*, 2011), but several others are available (Carrara *et al*, 2013). As the cost of conducting RNA-seq experiments is declining, the method quickly becomes more widely applicable.

Several challenges and limitations yet to be solved must be kept in mind when conducting and interpreting data from RNA-seq experiments. The RNA obtained reflects a defined moment in the cell cycle and is tissue specific. The data will represent RNA from all the cells analyzed, not just the tumor parenchyma cells. Also, no consensus exists as to which tissue is appropriate as “normal” tissue for comparison. In addition, all the steps in the experimental procedure are susceptible to errors.

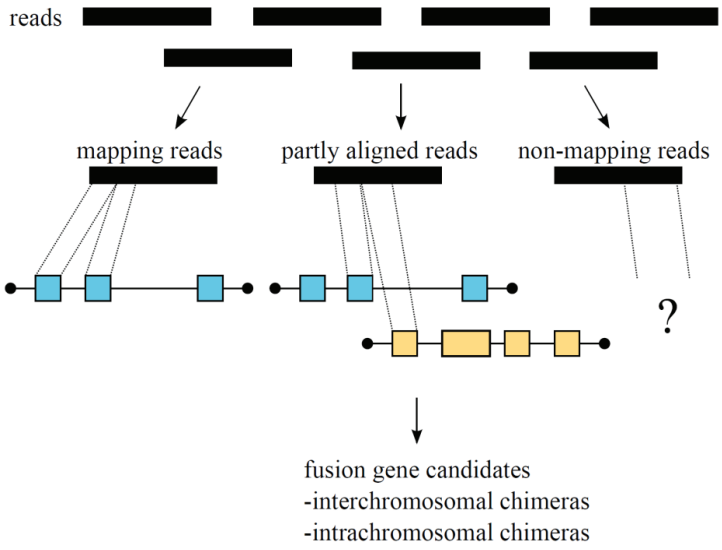


Figure 5: Fusion gene discovery from RNA-seq datasets. In principle, three different possibilities for alignment exist when the reads obtained by sequencing are compared to the reference genome: mapping reads (1), partly aligned reads (2) and non-mapping reads (3). Potential fusion genes are found in the second group, but additional filtering and validation are necessary for fusion gene detection. Modified from Maher et al (2009)

Aims

General aims

The study of chromosome changes in cancer cells has since the introduction of banding techniques in the 1970s (Caspersson *et al*, 1968, Smeets, 2004) provided many insights into the pathogenetic processes underlying tumorigenesis. Through persistent characterization of the breakpoints of cancer-specific chromosomal rearrangements, numerous new genes involved in human malignancies have been identified (Mitelman *et al*, 2007). At times, the complexity of acquired chromosomal changes is almost overwhelming, especially tumors of epithelial origin tend to display complex karyotypes (Heim and Mitelman, 2009). However, using the screening technique provided by cytogenetics, one can occasionally identify relatively simple but nonrandomly occurring chromosomal changes which are likely to influence cancer development. Further characterization of these changes using molecular cytogenetic and strictly molecular methods may reveal further molecular details. The translational strategy combining research efforts applied to acquired chromosomal rearrangements detected during diagnostic analyses of neoplastic cells has proven successful in detecting new cancer-specific chromosomal aberrations. From the regular diagnostic services provided by the Section for Cancer Cytogenetics, we have exclusive access to a large number of karyotypically characterized tumor samples of different origin, samples that show a variety of chromosomal aberrations.

The aim of the present thesis was to examine in detail such balanced chromosomal rearrangements detected in neoplastic cells and further characterize their consequences at the genomic level. Our aim was not to screen all patients within a disease entity, but rather to describe the pathogenetic events resulting from a few selected chromosomal aberrations that probably define a small subset of patients, and thereby elucidate an alternative way to cancer development. For that purpose we applied FISH and PCR based methods as well as transcriptome sequencing in the last two papers. These methods were used in different combinations to gain a complete overview of the genomic rearrangements and to identify their products at the RNA level. By means of

this multimodal approach, we could circumvent some of the limitations particular to each technique.

Today, most tumor classification systems are based on the morphologic appearance of the tumor cells and their tissue organization as it appears through the microscope. We hope that our findings will help identify additional diagnostic markers that can distinguish tumors based on their acquired mutations, as such a classification may reflect important tumor characteristics better than do their histological appearance (Brandal *et al*, 2010). Eventually, our hope is that a more profound understanding of disease-causing mutations will reveal new targets for specific treatments.

Additional specific aims and the theoretical rationale for each case studied are described in the paragraph dealing with each paper.

Specific aims

Paper I

Cytogenetic findings are crucial both for the diagnosis and prognosis of ALL patients (Group Francais de Cytogenetique Hematologique, 1996, Faderl *et al*, 1998, Harrison, 2009). The great majority of patients with this disease have cytogenetic abnormalities in their bone marrow cells, with detailed analysis that includes IP-FISH examination having demonstrated aberrations in up to 90% of newly diagnosed ALL cases (Harrison *et al*, 2005, Moorman *et al*, 2007). Up to 30% of ALLs have cytogenetic changes leading to fusion genes (Mitelman *et al*, 2007) with the type of fusion varying across age groups. In childhood ALL, the most common rearrangements are hyperdiploidy, followed by a t(12;21) recombining the genes *ETV6* and *RUNX1* (Moorman, 2012).

Prognosis in ALL is highly dependent on the leukemic karyotype at disease debut. An example is the group of patients with a high-hyperdiploid karyotype which corresponds to a favorable risk profile with nearly all patients surviving if treated adequately (Harrison, 2011). In contrast, patients with a low-hypodiploid karyotype have a far worse prognosis, one study measured event-free survival after three years of follow-up to only 29% in this group (Harrison *et al*, 2004). Examples of prognostic differences also exist among patients with structural rearrangements; patients with t(12;21) have an overall favorable prognosis while t(9;22) is associated with a poor prognosis necessitating allogeneic stem-cell transplantation in the first complete remission for these patients (Pui and Evans, 2006). Several other groups based on recurrent aberrations are acknowledged, but because of few reported cases, many of the cytogenetic aberrations in ALL are of unknown prognostic significance.

A 19-year-old patient diagnosed with B-precursor ALL was treated at our hospital. The cytogenetic analysis of his bone marrow aspirate revealed a 12;17-translocation as the sole karyotypic abnormality in the leukemic cells. The t(12;17)(p13;q12) is rare and of uncertain prognostic significance, but the reported data suggest a relatively good prognosis. Prior to our study, only 25 patients with this translocation had been reported, and in six of these, a fusion gene *TAF15-ZNF384* had been identified

(Martini *et al*, 2002, La Starza *et al*, 2005). Our aim was to investigate in molecular detail the consequences of the t(12;17). Were the same genes involved in our case? If so, was the breakpoint identical at the molecular level, i.e., were the same exons or introns involved? Was the fusion gene translated to an RNA transcript? Were there any cryptic rearrangements involved? Furthermore, we identified another ALL patient in our archive with a complex karyotype which included a similar t(12;17). We therefore examined also this second translocation addressing the questions mentioned above.

Paper II

AML is the most thoroughly studied malignancy in terms of cytogenetic aberrations with 16.998 cases (as of August 2013) reported in the Mitelman Database (Mitelman *et al*, 2013). The number of identified fusion genes in this disease is high and increasing. Roughly 60 % of patients with *de novo* AML display cytogenetic changes detectable by chromosome banding methods (Grimwade *et al*, 2010). In one series of young adult patients, the most common aberrations were trisomy 8, monosomy 7, t(8;21), and t(15;17) (Grimwade *et al*, 2010). As in ALL, the frequency of the aberrations depends on which age groups are examined. One of the characteristics of AML is that it is a heterogeneous disease. Numerous balanced translocations in addition to the ones mentioned above have been detected, only some of which carry a known diagnostic or prognostic impact (Heim and Mitelman, 2009). In addition to recognized cytogenetic subgroups, a panel of genes is found to be associated with prognosis if altered (for overview, see Gianfelici *et al* (Gianfelici *et al*, 2012)).

A 28-year-old patient with AML was admitted to his local hospital, and a bone marrow aspirate was sent to our laboratory for cytogenetic examination. The G-banding analysis revealed a t(12;13)(q14;q31) as the sole karyotypic change. We wanted to characterize the breakpoints in detail and determine the underlying gene-level changes. Which were the genes involved? Was there a fusion gene resulting from the rearrangement or is a deregulation of gene(s) the result of the translocation? What was the translocation product(s) at the RNA level? Was one of the well known cancer-related genes *MDM2* and/or *HMGA2,* both of which map to 12q14, involved? Or were

genes mapping to the 13q31 area crucially involved, for example the *GPC5* gene known to be amplified in selected cases of rhabdomyosarcoma (Williamson *et al*, 2007)?

Paper III

Mesenchymal chondrosarcoma (MC) is a rare subtype of chondrosarcoma (<3% of primary chondrosarcomas) (Fletcher *et al*, 2013). Most patients are middle-aged and most tumors arise in the craniofacial bones, the ribs or the vertebrae. A biphasic pattern with areas of round primitive mesenchymal cells interrupted by chondroid elements characterizes the tumor histologically (Nakashima *et al*, 1986).

Knowledge about the pathogenetic factors responsible for MC is limited, but a fusion gene identified in a majority of investigated cases was recently identified (Wang *et al*, 2012). The fusion joins the DNA binding domain of the *HEY1* gene to the C-terminal transcriptional activation domain of *NCOA2*, and it is the fourth fusion gene involving *NCOA2* identified in human cancer (Wang *et al*, 2012). The fusion was identified using a genome-wide exon based expression array and no karyotypic information about the tumors was provided. The authors suggested the rearrangement to be caused by a submicroscopic deletion on chromosome 8.

No consistent cytogenetic findings were described in the ten karyotyped cases of MC in the Mitelman database (Mitelman *et al*, 2013). Most karyotypes were complex. In two tumors, a Robertsonian translocation between chromosomes 13 and 21 was detected in addition to loss of all or a portion of chromosomes 8 and 20 and gain of all or a portion of chromosome 12 (Naumann *et al*, 2002). Only two previous cases with a simple karyotype were reported, one with trisomy 8 as the sole aberration (47,XX,+8) (Gatter *et al*, 2005), the other with a balanced t(4;19) (46,XY,t(4;19)(q35;q13)) (Richkind *et al*, 1996).

We analyzed tumor cells from a 63-year-old woman with an MC harboring a 1;5-translocation as the sole aberration identified at the cytogenetic level. Since the translocation was present in all cells cytogenetically analyzed, we karyotyped also cells from a peripheral blood sample to rule out the possibility of a rare constitutional

variant. Once we had established that the 1;5-translocation was indeed tumor specific, we decided to characterize this new aberration at the molecular level. Did the MC under examination perhaps harbor a submicroscopic deletion of chromosome 8 leading to the *HEY1-NCOA2*, the fusion gene published as typical of MC while we were working on our case? If so, would the putative deletion lead to the same fusion that had already been reported or would it result in a variant? If not, did the present tumor rearrangement lead to a totally different fusion gene? If a novel fusion existed, what was its result at the molecular level? What was the prevalence of this/these fusion genes in other MCs?

Paper IV

Pure erythroid leukemia is characterized by neoplastic proliferation of cells resembling proerythroblasts (Swerdlow, 2008). The disease is extremely rare and is recognized as a diagnostic entity together with erythroleukemia (erythroid/myeloid) under the term "acute erythroid leukemias" in the WHO classification system (Swerdlow, 2008). Differential diagnosis against megakaryoblastic leukemia may be challenging, especially in cases of pure erythroid leukemia without erythroid maturation. There is sometimes gradual progression from myelodysplasia with erythroid dysplasia to pure erythroid leukemia, and a diagnostic marker is needed (Domingo-Claros *et al*, 2002). The patients often follow a rapid clinical course even with the best current treatment.

Previous cytogenetic studies of patients with erythroleukemia have displayed numerical and structural aberrations resembling those seen in AML in general, especially trisomy 8 and structural changes leading to loss of material from chromosomes 5 and 7 (Mitelman *et al*, 2013).

An 18-month-old boy was admitted to our hospital and diagnosed with pure erythroid leukemia. G-banding analysis showed a karyotype interpreted as 46,XY,der(1)t(1;1)(p31;q21),t(1;16)(p31;q24). Extensive FISH studies revealed that the *CBFA2T3* gene was rearranged in the 1;16-translocation (Micci *et al*, 2011). However, because the material available for molecular analysis was degraded, we were not able to determine the leukemogenic mechanism whether it be the generation of a fusion gene with *CBFA2T3* as one of the participants or loss of tumor suppressor

activity, in which case *KANK1*or *L1TD1* could be the target since both these genes were found to be homozygously deleted. Furthermore, it looked as if the 1;16-translocation is recurrent in pure erythroid leukemia since two additional cases have been reported with the same rearrangement (Koller *et al*, 1989, Castaneda *et al*, 1991). A few months after the first publication (Micci et al., 2011), we got hold of an additional 2 ml of coagulated blood with 3% leukemic cells from the patient. Since the technology in the field of sequencing had developed so fast, we decided to extract RNA from the sample and try high-throughput sequencing analysis.

Our aim was now to characterize the t(1;16) at the molecular level. Would it be possible to identify the putative fusion gene using high throughput RNA sequencing working from a specimen with such a small number of abnormal cells? Was there a fusion gene between chromosomes 1 and 16? If so, which were the fusion partners, and could they be identified by PCR? What would be the possible mechanisms by which the chimeric protein executes its oncogenic potential? Since the FISH studies had identified a more complex karyotype, described as 46,XY,der(1)t(1;1)(p31;q21), del(1)(p11p31),der(16)t(1;16)(p31;q24), we were also wondering if the other described chromosomal breakpoints in the karyotype generated fusion genes.

Results in brief

Paper I

The first description of the translocation found in this case was t(12;17)(p13;q21), and although the most likely fusion gene was the known but rare *TAF15-ZNF384* resulting from a t(12;17)(p13;q12), there were also several other potential candidate genes. *ETV6* is located on 12p13 and is one of the most promiscuous gene partners in leukemia. To date, >30 different *ETV6* partner genes have been identified (De Braekeleer *et al*, 2012, de Braekeleer *et al*, 2013). We used a commercially available *ETV6/RUNX1* dual-color translocation probe to rule out the involvement of *ETV6*. We sat up a FISH-BAC assay to narrow down the breakpoints further. BAC RP11-151M4 overlapping eight genes including the *ZNF384* gene gave a split signal when hybridized to patient metaphase cells. We then used fosmid clones to determine the breakpoint with greater precision. This latter approach did not succeed, probably due to repeated hybridizations to the same slide, and gave too weak signals to interpret. Meanwhile, we identified a splitting BAC overlapping the *TAF15* gene on 17q12. Our results so far pointed to the fusion gene *TAF15-ZNF384* and the karyotype was reinterpreted as t(12;17)(p13;q12). Both BACs covered an area of about 150 kb. This is a large area to analyze at the molecular level and would require many PCR reactions since we did not know where the precise breakpoint was. We assumed that the breakpoints were the same as those detected by Martini et al. (2002), and therefore use the same primers. We successfully amplified a *TAF15-ZNF384* fragment.

We then wanted to investigate whether the *TAF15-ZNF384* was present also in the second ALL patient identified through a search of our patient archives. FISH and PCR experiments confirmed the presence of a *TAF15-ZNF384* fusion from DNA extracted from the leukemic cells also in that case.

Sanger sequencing of the PCR products from the two cases demonstrated that the break in *TAF15* had occurred within intron nine in the first patient and within intron six in the second patient. In the *ZNF384* gene, the break occurred within intron two in

both cases. Only in patient 1 was RNA available, and the transcript could be confirmed.

Paper II

To characterize the breakpoint position of the t(12;13) we began with FISH using specific BAC clones. Since we had no putative fusion genes for this case, we scattered the probes in chromosomal bands 12q14-q21, an area of approximately 20 Mb (UCSC Genome Browser (Kent *et al*, 2002). In total, 36 BAC probes were tested before two splitting probes (one on the derivative chromosome 12 and one on the derivative chromosome 13) were identified. On chromosomal band 12q14 a split signal was seen using a BAC which overlaps the *HMGA2* gene. Fosmid clones were used to narrow down the breakpoint, and we were able to map it to the 3`end of the gene. The breakpoint on chromosomal band 13q31 was found to be in a region where no known genes are located. A fusion gene was therefore an unlikely consequence of this translocation. Attempts to amplify the breakpoint region on genomic DNA were unsuccessful, so in order to describe the precise rearrangement, we ran a RACE PCR amplifying the 5`end of *HMGA2.* We identified a truncated form of *HMGA2* by sequencing the resulting transcript. To demonstrate that the truncated transcript had consequences also at the RNA and protein level, we wanted to quantify the expression of *HMGA2* by qPCR and immunohistochemistry. *HMGA2* expression was considerable increased compared to a control group of AML patients assed by qPCR. Increased expression of HMGA2 was also demonstrated at the protein level, as we by immunohistochemistry saw specific staining of the blast cell nuclei.

Paper III

Having identified an MC patient with a tumor-specific t(1;5), we searched for additional patients within the Sarcoma Registry at the Norwegian Radium Hospital to find patients with the same diagnosis. From 1986 to 2010 only 3 such patients were identified, underlining the rarity of these tumors. We got hold of frozen material from all three tumors, and run a reverse transcriptase PCR aiming to amplify the *HEY1-NCOA2* fusion using the same primers as used by Wang *et al* (2012). A PCR fragment

was amplified in all but the t(1;5) case, and subsequent Sanger sequencing confirmed the presence of a *HEY1-NCOA2* transcript in the other tumors.

The splitting BAC on 5q33 overlapped four genes. Using fosmid clones we narrowed the area further, to the genes *PDGFRβ* and *CDX1*. Transcriptome sequencing was performed and the "deFuse" algorithm (McPherson *et al*, 2011) was used to align sequences and predict potential fusion genes.

The list of predicted fusion genes exceeded 100. We focused on putative fusions between chromosomes 1 and 5, as we knew these two chromosomes were rearranged. A putative *IRF2BP2-CDX1*transcript topped the list of potential fusions when ranked by split count, i.e., the number of reads supporting the prediction. The *IRF2BP2-CDX1* was validated using primers outside the predicted transcript, and was also detected on genomic DNA extracted from the tumor. The fusion was not identified in any of the other described cases.

Paper IV

The blood sample from the patient harboring the t(1;16) contained 3% leukemic cells. RNA was extracted and sent for sequencing. A total of 107 million reads were obtained. The filtered data were run through the Fusion Map software (Ge *et al*, 2011), and a list with about 500 putative fusion genes was identified. As number ten in this list ordered by the number of reads supporting the predicted fusion, a fusion between the genes *CBFA2T3* and *NFIA* was identified. These genes map to chromosome bands 16q24 and 1p31, respectively, which fits well with the described karyotype and the previous FISH results. The *NFIA/CBFA2T3* fusion was verified by PCR and subsequent Sanger sequencing. As the FISH studies had revealed that the t(1;16) was in fact more complex than originally thought, *in silico* analysis was performed of the sequences that involved chromosome 1. All predicted transcripts contained repetitive sequences or sequences with a high sequence identity to other genes, making them unlikely fusion partners.

The fusion gene gave an open reading frame and was predicted to encode a protein consisting of 811 amino acids, combining 208 amino-acid residues from *NFIA* and 603 residues from *CBFA2T3*.

Discussion

Methodological considerations

In this section, I shall discuss the study design of the subprojects that form the basis for this thesis in addition to some aspects of the methods applied. General aspects of each of the utilized method were discussed in the Introduction.

Study design

In all the studies presented in this thesis, only one or a few patient samples were examined. The selected cases were those with rare, cancer-specific cytogenetic aberrations, all identified through our diagnostic service. It is unquestionable that case reports and small series cannot replace well-designed, statistically-oriented studies when focusing on a particular therapy or the efficiency of a certain diagnostic test. However, they can be equally important when aimed in the right direction (Vandenbroucke, 2001). Case studies have historically provided significant advances also in cancer research. The famous *BCR-ABL* story started with the identification of a minute chromosome in cells from two patients with chronic leukemia (Nowell and Hungerford, 1960a). Nowadays, single-patient studies are the first to be presented when a new technology is available, but still too expensive and time-consuming for the general research community. One such study was published in 2008 when the entire genome of a cytogenetically normal AML patient was sequenced (Ley *et al*, 2008). Unusual effects of a drug be it adverse or beneficial, are also often first reported as a case report. In other words, case studies are hypothesis-generating for forthcoming studies in the field, but cannot have statistical value or be directly implemented in evidence-based medicine.

We used G-banding of chromosomal preparations as an initial screening technique to identify clonal chromosomal aberrations. Cytogenetic analysis is a robust technique that has survived the changing trends in cancer research since the correct chromosome number was identified in 1956 (Tjio and Levan, 1956). It offers an overview of the cell genome which can be directly visualized down the microscope without the need for

complex algorithms. Some limitations of the method must be kept in mind: dividing cells are required for cytogenetic analysis, and some tumor cells do not divide *in vitro,* resulting in a false negative result (i.e., a normal karyotype). This has turned out to apply mostly to tumors of epithelial origin (Edwards, 2010). The number of aberrant cells represents the neoplastic cells' ability to divide *in vitro* rather than *in vivo*, and so analysis of G-banded chromosomes is not ideal to assess the relative size of a given tumor or leukemia clone. Also, chromosome morphology may be too poor to identify the karyotypic abnormalities (Mitelman *et al*, 2007). The cases singled out for scrutiny in this thesis had cells that were able to divide *in vitro.* Cases with balanced translocations (with the exception of the case studied in paper IV and of one subproject for which a complete molecular characterization was not obtained; see Table 2) were chosen because these changes most probably are important for the neoplastic transformation.

Working with limited material

Studying cases identified through the diagnostic service had some drawbacks. Material available for analysis was often sparse and had not been optimally collected for research purposes. We had no means of controlling how much tissue was left for analysis or for how long the tissue was stored at room temperature before freezing (something that affects the quality of extracted nucleic acids). The following example demonstrates this limitation: in paper II, where good quality RNA was available from the malignant cells, we were able to run a 3`RACE PCR after the FISH results indicated involvement of the *HMGA2* gene. In contrast, the same technique was not feasible in the study leading to paper III due to degraded RNA.

After karyotyping, we applied FISH-based methods whenever feasible. To meet the challenge of sparse material, we tried to re-use previously G-banded metaphase spreads for FISH analysis, an approach that has been successfully applied in our laboratory before (Micci *et al*, 2002). The advantage is that the identity of the chromosomes studied is known beforehand, something that makes the interpretation of weak signals easier. Slides can thus be used for several rounds of hybridization. The identification of the fusion gene *TAF15-ZNF384* in paper I was made possible by this

approach when only two abnormal metaphases were available for analysis. However, more material for analysis would have been necessary had we not suspected that these two genes were the ones involved. The downside of re-using material is the harshness of the destaining protocol which wears down the biological material. Sequential hybridizations first remove proteins, but repeated hybridizations also reduce the DNA content of the chromosomes, making signal intensity weaker. Repeated hybridizations may also increase unspecific hybridization because of remaining debris from previous procedures.

Even with these efforts to save and re-use material, four additional projects planned to be a part of this thesis had to be discontinued because of lack of material before a complete molecular characterization of the cytogenetic aberration was achieved. (Table 2 provides an overview). Three of these cases were sarcomas characterized by apparently balanced translocations. The main problem working with these cases was that so few metaphases were available for FISH studies. One example was the case of a chondrosarcoma showing a 5;10-translocation as the sole karyotypic change. In the 20 slides dropped only four abnormal metaphases were identified. Neither in this case nor in any of the other sarcoma cases was the RNA quality good enough to perform transcriptome sequencing.

Malignancy	**Chromosomal aberration**
Chondrosarcoma	t(5;10)(q13;q26)
Liposarcoma	t(3;9)(p25;q22)
Sarcoma	t(2;17)(q23;q23)
Leukemia	der(9)t(9;12)(p23;q13)*

Table 2: Projects that had to be ended because of lacking material.*Part of complex karyotype.

Locus-specific FISH

Locus-specific FISH was used in papers I-III. This technique was crucial for our experiments before we implemented transcriptome sequencing as a tool. To determine a breakpoint position, clear and recognizable signals are necessary, as the staining used with FISH (DAPI) only displays the chromosome size and breakpoint position. From time to time we experienced background hybridization to be a problem with locus-specific FISH, especially when using old and destained slides (Figure 6). There were three main issues or problems with the FISH protocol: unspecific hybridization of the probe to the chromosomes, insufficient pretreatment of the slides, and poor probe quality. Addressing the first issue, we assessed the necessary amount of Cot-DNA. Cot-DNA is added to FISH experiments because it hybridizes to repetitive sequences in the genome, preventing false signals caused by unspecific binding (Fan, 2002). However, changing the concentration of Cot-DNA did not alter background signals. Next, we modified the pre and post hybridization washes, but with no improvement of results. Third, the prepared BAC probes were evaluated. When extracted from the bacterial cell, the BAC DNA is mixed with bacterial DNA (which exceeds BAC DNA by far), and we applied a different protocol to improve the BAC DNA output. Unfortunately, no consistent results favoring one or the other modification were obtained and troubleshooting of such a comprehensive protocol turned out to be a difficult task. Most likely, physical and/or chemical properties of the probe in combination with the slides contributed to the background signals.

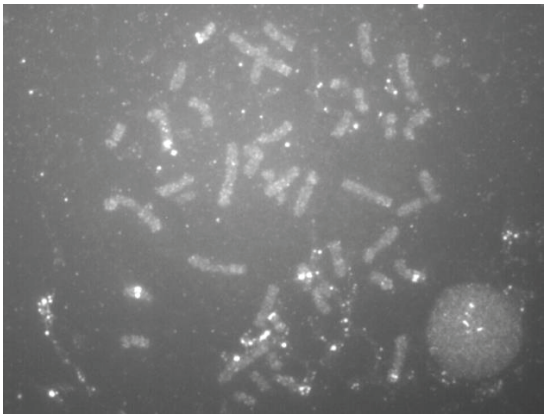


Figure 6: Hybridization of this probe resulted in both chromosomal and extrachromosomal signals, making interpretation difficult

PCR-based methods

We used PCR-based methods to characterize the breakpoint area defined by FISH. Our choice of method in each case was highly dependent on the amount of available material.

The genes involved in chromosomal rearrangements may contain large introns. Trying to amplify a hybrid gene from genomic DNA can therefore result in long PCR fragments, sometimes longer than the capacity of the enzymes used for the PCR reaction. We met this challenge by optimizing the PCR protocol for amplifying long fragments (LD-PCR). This variant of PCR is more susceptible to false positive PCR products because of its long extension time. Therefore, if a band appeared as a result of LD-PCR on genomic DNA, we ran more reactions involving neighboring exons. Whenever possible we preferred RNA as a template for RT-PCR. The "trial and error" approach we used when we had identified two putative genes does not work if unexpected submicroscopic rearrangements are present at the breakpoint position. As an example, the *IRF2BP2-CDX1* fusion gene (paper III) would not have been easy to identify using this approach.

We used qPCR in paper II to assess the expression of *HMGA2*. As discussed in the introduction, there are many pitfalls to qPCR, one being that it is difficult to select an appropriate control group. Not much is known about *HMGA2* in hematological malignancies, but it seems that neoplasia-associated rearrangements involving chromosomal bands 12q13-15 lead to inappropriate expression of the protein (Fusco and Fedele, 2007). We chose to include AML patients with a common cytogenetic aberration, trisomy 8 (Heim and Mitelman, 2009) but without any visible changes of chromosome 12 and a "normal" bone marrow sample as a control group. We figured that comparison only with bone marrow from healthy donors could give an incorrect estimate of the gene expression as several genes may be aberrantly expressed in AML in general. We identified two AML patients with trisomy 8 as the only cytogenetic aberration, and one patient with monosomy 7 in addition to trisomy 8. A larger control group would have given us a more robust result, but the set up for this study limited us in this regard. As an internal control we chose the *GUSB* gene, a commonly used gene

for endogenous control in qPCR studies, and one of the genes suggested by Beillard and colleagues as suitable for detecting minimal residual disease in leukemia (Beillard *et al*, 2003).

Identifying fusion genes by transcriptome sequencing

The traditional way to identify fusion genes, the use of locus-specific FISH followed by PCR-based methods, was applied in papers I and II. The drawback of this approach is that is can be labor-intensive and requires good quality metaphases in sufficient number. In papers III and IV we used RNA sequencing in combination with G-banding. This rather new method for fusion gene detection allowed us to target our analysis to the chromosomal aberration described cytogenetically. Identification of fusion genes in RNA-seq datasets was first reported in 2009 (Maher *et al*, 2009). Tanas and colleagues were the first group to implement cytogenetic methods together with transcriptome sequencing, thus identifying a fusion gene in epithelioid hemangioendothelioma (a rare vascular sarcoma) (Tanas *et al*, 2011). Analysis of transcriptome data can be challenging, but we avoid some problems by combining cytogenetics and sequencing. Several algorithms have been proposed for fusion gene detection in RNA-seq datasets, all with different biases that must be taken into consideration (Carrara *et al*, 2013). Some algorithms apply a more stringent filtering than others, aiming to reduce the number of false positive predictions. The risk with this approach is that also true transcripts may be eliminated. FusionMap (applied in paper IV) uses less stringent filtering than deFuse (paper III) and requires less computational recourses. When a putative fusion region is already identified through cytogenetic analysis, the problem with false positive predictions becomes less challenging.

Detection of fusion genes resulting from cryptic genomic aberrations is now possible because of advances in genomic technologies. Through analysis of data from array-based studies, several investigators have reported novel and recurrent gene fusions (Tomlins *et al*, 2005, Wang *et al*, 2012, Giacomini *et al*, 2013). Although such studies are said to be unbiased, informed choices must nonetheless be made when analyzing the data. The large datasets generated compel researchers to filter the data in some

manner, and often genes not previously linked to cancer are excluded from further analysis, risking elimination of pathogenically important genes.

These different methods for identifying altered genes complement each other. The *HEY1-NCOA2* recently detected in MCs would probably not have been identified cytogenetically, as it results from an interstitial deletion, while the t(1;5) if present in a subset of MC seems to be too rare to be identified in studies of only a few patients; it is probably a rare aberration in a rare cancer.

Biological considerations

In this thesis, three chromosomal rearrangements in acute leukemia and one rearrangement in an MC were investigated in molecular detail. The aims of the studies were descriptive, i.e., we wanted to characterize the rearranged gene(s) at the genome level. We also assessed the consequences at the RNA level, whenever possible.

Why study rare cancers?

Despite the fact that most research efforts are directed towards common malignancies such as cancers of the prostate, breast, lung, and colon (Cancer Registry of Norway, 2013), we decided to study samples from some rare tumor types. We see several reasons why also rare cancers may be worth studying. First, in some perspectives the division between rare and common cancers is artificial. Molecular profiling of common tumors have demonstrated that they are the phenotypic outcome of numerous and different pathogenetic processes (Munoz and Kurzrock, 2012). Based on the notion that there are fewer pathways that lead to the development of rare cancers (Braiteh and Kurzrock, 2007), rare cancers may be more homogenous, and therefore easier to study. In fact, most of the successful therapies recently developed are directed towards rare cancers (Braiteh and Kurzrock, 2007). Occasionally, discoveries in rare cancer types are useful also in other, more common cancers. One example is found in a subgroup of lung cancers; activation of the *ALK* gene through a gene fusion was first identified in anaplastic large-cell lymphoma (Kutok and Aster, 2002). The same fusion gene has subsequently been identified in about 4% of non small cell lung carcinomas

(Soda *et al*, 2007), and targeted therapy inhibiting *ALK* is now implemented in the treatment of these patients (Kwak *et al*, 2010).

Why study balanced translocations?

In the "cytogenetic dictionary", chromosomal aberrations are divided into primary and secondary aberrations, in addition to a third type, called cytogenetic noise (Heim and Mitelman, 1989). The primary aberrations are thought to be the tumor initiating events, i.e., necessary for cancer development. These aberrations are often found alone and are correlated with a distinct tumor type. Secondary aberrations, on the other hand, rarely exist alone and are believed to occur in cancer cells already harboring primary aberrations. They are important in tumor progression, and are probably selected in a Darwinian fashion. Cytogenetic noise is a term for non-clonal aberrations that occur because of genomic instability (Heim and Mitelman, 1989).

The division into "drivers" and "passengers" is another way of categorizing mutations in cancer used frequently today (Stratton *et al*, 2009). "Driver" mutations are those thought to be causally implicated in oncogenesis. Passenger mutations are not believed to confer a growth advantage on the cell carrying them, and therefore do not contribute to cancer development (Stratton *et al*, 2009).

Translation between these two set of terms is not always easy. Cytogenetic aberrations may refer to loss or gain of a whole or parts of a chromosome, while the drivers and passengers more often refer to single base substitutions within genes. On other occasions the translation is more straightforward. Balanced translocations, for example, are regarded as primary cytogenetic aberrations when identified as a sole cytogenetic change, and the fusion genes that such aberrations often form are "drivers" of the oncogenic process.

We work from the premise that the simple chromosome aberrations such as those examined in this thesis represent primary abnormalities and/or driver mutations, something that makes them all the more interesting to study. Although we cannot be certain about their contribution to oncogenesis, several considerations, some described below, support this hypothesis.

Genes involved in transcriptional regulation and nucleic acid binding are overrepresented among genes involved in human malignancies (Furney *et al*, 2006). Fusion gene partners are, according to recent studies, either genes involved in transcriptional control or they act as tyrosine kinases (Mitelman *et al*, 2004). All seven genes identified through the chromosomal translocations studied by us as part of this thesis influence the transcription process. *ZNF384*, *CDX1*, *HMGA2* and *NFIA* are all transcription factors and bind directly to DNA (Vaquerizas *et al*, 2009). *TAF15* interacts with transcription factor *TBP*, *IRF2BP2* interacts with transcription factor *IRF,* and *CBFA2T3* acts as a transcriptional repressor (Kochetkova *et al*, 2002).

The fusion transcripts found in leukemias and sarcomas are often composed of parts of genes involved in gene regulation or chromatin-modification (Edwards, 2010). The fact that all the four fusion genes identified (*TAF15-ZNF384, HEY1-NCOA2, IRF2BP2-CDX1,* and *NFIA-CBFA2T3*) join two gene-regulating genes therefore supports a functional role in oncogenesis. Also, structural similarities to already well-known fusion genes strengthen the probability of a role in cancer development. This is exemplified by the *TAF15* gene, first identified in 2002 (Martini *et al*, 2002). TAF15 belongs to the TET protein family. It is structurally similar to another, more famous member of the same family, specifically the Ewing sarcoma protein known from the childhood malignancy carrying the same name (Tan and Manley, 2009). Not surprisingly, *EWSR1* was demonstrated to form pair with *ZNF384* in other leukemia patients (Martini *et al*, 2002). Such parallels can also be drawn for the *CDX1* gene that forms the 3`end of the *IRF2BP2-CDX1*gene, although less powerful. *CDX2*, closely related to *CDX1,* has been detected as a part of a fusion gene in leukemia (Chase *et al*, 1999). Structural similarities also exist between *NFIA-CBFA2T3* and other, more well-known fusion genes, namely the *RUNX1-CBFA2T3* resulting from a t(16;21) (Gamou *et al*, 1998)

Two of the fusion genes identified in this thesis are already known to be recurrent (*TAF15-ZNF384, HEY1-NCOA2*), and this supports their role in oncogenesis. As mentioned above, the *NFIA-CBFA2T3* fusion is also likely to be recurrent, since a description of two patients with erythroleukemia and a t(1;16)(p31;q24) in their

karyotype was reported in the Mitelman database (Koller *et al*, 1989, Castaneda *et al*, 1991).

In paper II we investigated a novel t(12;13) identified in a patient with AML. Instead of finding a new fusion gene as we had expected, we mapped the breakpoint in 12q13 to inside the *HMGA2* gene. No gene resides at the breakpoint position identified on chromosome 13q. This mechanism of gene alteration is not the typical consequence of a balanced translocation as translocations usually cause either a fusion gene or deregulation of one gene through juxtaposition with regulatory sequences from another gene (see Introduction). Deregulation of *HMGA2* through 12q13~15 rearrangements is nevertheless well-known in benign mesenchymal tumors, lipomas in particular (Fusco and Fedele, 2007). The growth benefit of these tumors is suggested to be caused by a loss of a repressive miRNA targeting the 3` end of *HMGA2* (Mayr *et al*, 2007). When overexpression of *HMGA2* and the closely related *HMGA1* is detected in malignant tumors, it seems to be correlated with a highly malignant phenotype and reduced survival (Fusco and Fedele, 2007). Knowledge about *HMGA2*'s role in hematological malignancies is limited, but the mechanism of deregulation seems to be similar to that observed in the case reported by us; *HMGA2* is activated by a translocation involving 12q, and no expressed fusion genes are identified. Hematological malignancies of different types have been associated with *HMGA2* deregulation. (Table 3 provides an overview of the reported cases).

Author	Disease	Abbreviation	Results	Other mutation
(Aliano *et al*, 2007)	PV	t(12;21)(q14;q22)	FISH+ *HMGA2* expression	*JAK2*
(Andrieux *et al*, 2004)	MMM	t(4;12)(q33;q15-q21) t(5;12)(p14;q15-q21)	FISH+ *HMGA2* expression Truncated *HMGA2* (ex 1-4)	NR
(Inoue *et al*, 2006)	PNH (benign)	t(12;12)(q13;q15)	FISH+ *HMGA2* expression (ex5/ 3`UTR)	*PIGA*
(Kottickal *et al*, 1998)	AML AML	t(5;12)(q10;q10) t(7;12)(p22;q13), t(12;13)(q10;q10)	Isoforms of *HMGA2* detected	NR
(Meyer *et al*, 2007)	CML(14 pt)	t(9;22) (13 pt) Normal (1 pt)	*HMGA2* expression correlated to WBC count	*BCR-ABL*
(Odero *et al*, 2005)	MDS (RAEB1) MDS (RAEB2) MDS (RAEB1) MDS/MPD MDS/MPD (aCML) Secondary MDS (RAEB1)	t(7;12)(p12;q13) t(12;14)(q13;q31) t(12;12)(p11;q13) t(12;20)(q15;q11.2) t(8;12)(q22;q13) t(11;12)(q23;q15)	FISH+ *HMGA2* expression Truncated *HMGA2* transcripts detected in two patients (ex 1-3 and ex 1-3 together with ex 1-2).	NR
(Pierantoni *et al*, 2003)	ALL	t(9;12)(p22;q14)	FISH+, Truncated *HMGA2* (ex 1-3)	NR
(Rommel *et al*, 1997)	CML (11 pt) AML (3 pt)	t(9;22) NR	*HMGA2* expression	*BCR-ABL*
(Santulli *et al*, 2000)	CLL	t(12;14)(q13;q32),t(9;13)(p13;q14)	FISH+, HMGA2 expression by immunohistochemistry	NR
(Storlazzi *et al*, 2006)	PV	t(3;12)(q26;q14)	FISH+, *HMGA2* (ex1-5)-*TNIK* detected but not expressed. *HMGA2* expression	NR

Table 3: Hematological neoplasias with *HMGA2* involvement. Abbreviations: PV: Polycytemia vera, MMM: Myelofibrosis with myeloid metaplasia, PNH: Paroxysmal nocturnal hematuria, MDS: Myelodysplastic syndrome, RAEB: Refractory anemia with excess blasts, MPD: Myeloproliferative disorder, ALL: Acute lymphoblastic leukemia, CML: Chronic myeloid leukemia, CLL: Chronic lymphocytic leukemia, FISH+: FISH results showing *HMGA2* involvement, ex: exon, NR: Not reported, WBC: White blood cells, pt: patient

Do the identified fusion genes define distinct disease subgroups?

Fusion genes are in principal tumor type specific, and therefore constitute a useful tool in diagnostics (Mitelman *et al*, 2007). This recognition makes fusion genes important for pathological diagnosis in a range of hematological and mesenchymal tumors.

However, for most tumors, such disease specific biomarkers are absent. The MCs were until recently such a tumor entity lacking a distinctive marker and the diagnosis depended solely on histological characteristics and immunostaining (Fletcher *et al*, 2013). The *HEY1-NCOA2* fusion gene (Wang *et al*, 2012), identified while we were working on paper III, now comes across as such a new and specific marker. So far, 29 cases have been analyzed, with the identification of the gene *HEY1-NCOA2* in 21 of 29 (72%) of the tested patients (Nakayama *et al*, 2012, Wang *et al*, 2012) and paper III. *IRF2BP2-CDX1* is the second fusion gene to be identified in this tumor type and may constitute an alternative pathogenetic pathway. However, since it so far is identified in only one tumor, the *IRF2BP2-CDX1* may be unique to these tumor cells. Also, when more than one fusion gene is identified in a sarcoma subtype, they usually activate the same oncogenic pathway (Tan and Manley, 2009). The rarity of MCs makes it difficult to investigate more tumors and thus clarify these issues.

Translocations involving the *ZNF384* gene on 12p13 (paper I) seem to define a subgroup of ALL patients (La Starza *et al*, 2005). In addition to the t(12;17), *ZNF384* is also involved in a t(12;22) and a t(12;19). The clinical characteristics of the patients reported by us in paper I are in accordance with previously reported cases. Patients are usually young adults with an initial good response to chemotherapy (La Starza *et al*, 2005). However, the total number of reported cases is too small to conclude as to the prognostic impact of these rearrangements.

The clinical characteristics of leukemia patients characterized cytogenetically by the t(1;16) examined in paper IV seem to correspond to a distinct clinical subtype. Three patients, including the one reported in paper IV, have so far been identified with a t(1;16)(p13;q24) and leukemic blasts of the erythroid lineage. They were all young boys and the outcome was poor.

Is one fusion gene enough to generate a malignant cell clone?

The oncogenic potential of the genes identified by us, relative to other possible mutations in the tumors examined, is difficult to assess. Most available data imply that more than one mutation is required to generate a malignancy (Mitelman *et al*, 2007).

On the other hand, *in vitro* studies of cells transfected with *EWSR1-FLI1*, the dominating fusion gene in Ewing sarcoma, showed the development of a tumor-like phenotype in spite of the fact that only a single oncogenic event had occurred (Riggi *et al*, 2005). The oncogenic contribution of fusion genes can also be studied in transgenic mouse models. Introduction of the *FUS-DDIT3* fusion gene specific for liposarcoma into the mouse genome resulted in the development of liposarcomas resembling those seen in humans (Perez-Losada *et al*, 2000). This fusion gene seems to require a susceptible cell type to cause tumor growth; although the chimeric protein was identified in several tissues, only from adipocytes did tumors develop. The *BCR-ABL* fusion transcript of CML also seems to be sufficient to cause disease; however, additional genetic events are strongly correlated with disease progression (Deininger *et al*, 2000). Data also suggest that a fusion gene has to be present "at the wrong place, at the wrong time" in order to cause disease; *BCR-ABL* transcripts can be detected at low levels in the blood from normal individuals (Bose *et al*, 1998).

In all the leukemia cases described in papers I, II and IV (except for the additional patient identified in our database), a scan for mutations of known leukemia-associated genes was performed at the Department of Pathology at our hospital. This was done as part of the diagnostic routine. In papers I and II, alterations of the *FLT3* gene were identified. Point mutations in this gene are associated with a poor prognosis in AML patients, and usually co-exist with other oncogenic mutations (Kottaridis *et al*, 2001, Gilliland and Griffin, 2002). Newer data obtained from sequencing studies demonstrate that a typical solid tumor harbors 33-66 mutations expected to alter protein expression, while the corresponding number for pediatric tumors and leukemias is 9.6 mutations (Vogelstein *et al*, 2013). Therefore, it is entirely possible, perhaps even likely, that the malignancies studied by us harbored also other, not recognized, mutations.

Consequences of gene alteration on protein expression

If a fusion gene/gene mutation is identified, it can in theory be translated into a protein with new and oncogenic capabilities, or it can result in deregulation of one of the implied genes, usually overexpression (Mitelman *et al*, 2007). Most genes

characterized in this thesis were examined both at the DNA (FISH, PCR on genomic DNA) and the RNA (RT-PCR, qPCR and RNA-seq) level. In papers II and III, we also wanted to assess whether the altered genes influenced protein expression. In paper II, immunohistochemistry targeting *HMGA2* demonstrated specific staining in blast nuclei. HMGA2 is not normally expressed in adult tissues (Fusco and Fedele, 2007) and detection by immunohistochemistry therefore implies abnormal expression of this protein. In paper III, CDX1 protein expression would be an indirect assessment for the translation of the *IRF2BP2-CDX1* fusion gene. CDX1 antibodies were commercially available, but repeated control experiments resulted in unspecific staining. Therefore, we were not able to evaluate if a fusion protein was formed using immunohistochemistry. Some assumptions about gene translation can still be made; using the "Open Reading Frame Finder" at NCBI both the novel fusion transcripts *IRF2BP2-CDX1* and *NFIA-CBFA2T3* were predicted to give an open reading frame (http://www.ncbi.nlm.nih.gov/projects/gorf/), supporting the notion that these fusion genes are translated into a protein.

The *TAF15-ZNF384* gene has been functionally evaluated before; the fusion showed transforming properties in a cell assay (Martini *et al*, 2002). From time to time, however, a balanced translocation can result in a truncated mRNA transcript, either due to a preterm stop codon in the coding sequence or a missense mutation resulting in an mRNA with altered protein-generating capabilities. This consequence can also be of pathogenetic importance if affecting a tumor suppressor gene. A wide range of methods exists to assess the functional significance of gene fusions. Some of these are exemplified in Table 4. Cell assays would be a natural starting point if one were to examine the fusion genes identified by us at the protein level.

Setting	Study subject	End point	Requirements	Example study
In vitro	Cell lines	Cell viability	Tumor cells/cell line expression the fusion gene in question	Giacomini *et al*, (2013)
		Cell proliferation		
		Cell invasion		
		Cell transforming properties		Maritini *et al*, (2002)
In vivo	Transgenic mice	Tumor formation/development of characteristic tumors	A viable transgenic mouse model	Perez-Losada *et al*, (2000)
		Disease-specific survival		
In vivo	Human beings (Phase I study)	Progression free survival	Patients harboring the fusion gene in question and a small molecule inhibitor targeting it	Kwak *et al*, (2010)

Table 4: Examples of methods for evaluation of the oncogenic contribution of a fusion gene

The structure of breakpoints and the fusion genes they cause

We have described in detail the breakpoints, the involved genes, and the fusion transcripts identified. By similar descriptions also in future cases, it can be clarified which parts of the involved genes account for the oncogenic potential. This knowledge can be essential if a potential interfering molecule is identified. One fusion gene can also exist in slightly different versions, i.e., involving different exons from the involved genes. This may again lead to fusion proteins with different transcriptional activities that again affect tumor phenotype, progression and prognosis (Zoubek *et al*, 1996, de Alava *et al*, 1998, Gonzalez *et al*, 2007). The *TAF15-ZNF384* fusion had been identified in two different versions before, one containing exons 1-9 of *TAF15* and the other exons 1-6, both fused to exon 3 of *ZNF384* (Martini *et al*, 2002). We confirmed these two fusion gene variants that exons 1-6 of *TAF15* are sufficient for development of the disease. Potential prognostic differences caused by these two

TAF15-ZNF384 transcripts are impossible to assess due to the low number of patients reported.

The breakpoint region of the *HEY1-NCOA2* fusion was identical in our study (paper III) and all other analyzed cases (Wang *et al*, 2012). Future analyses of the *IRF2BP2-CDX1* and *NFIA-CBFA2T3* genes will reveal if different variants exist for these fusions.

In paper III, we expected that a balanced translocation had occurred between chromosomes 1 and 5. However, due to the orientation of the involved genes (*IRF2BP2* and *CDX1*), some additional rearrangement must have taken place. We tried to investigate this issue further by locus-specific FISH overlapping the breakpoint position on chromosome band 1q42. We validated *IRF2BP2* as one of the fusion partners and expected a BAC probe overlapping *IRF2BP2* to yield a split signal. Instead, we saw two signals located on chromosome 1. We concluded that a submicroscopic rearrangement must have taken place, but were not able to characterize the rearrangement further.

Conclusions and future perspectives

In the papers presented in this thesis we have described the molecular consequences of four different cytogenetic aberrations in malignant cells. This has resulted in better characterization of two infrequent aberrations whose gene-level consequences were already known, and the detection of two new cancer-specific fusion genes.

The aim of cancer research should ultimately be to reduce the mortality and morbidity associated with cancer. There are many ways in which this goal can be attained. Our work may first have an impact diagnostically, as numerous genetic aberrations have been demonstrated to be disease specific (Mitelman *et al*, 2007). Improved diagnostics would contribute to better cancer treatment, by selection of the right patients for the right treatment. Stratifying patients on the basis of genetic aberrations is already implemented for the most common hematological neoplasias, and has resulted in differences in treatment protocols based on the known prognostic impact of certain aberrations (Swerdlow, 2008).

Our group has for many years studied such changes and the research continues. With the recent introduction of transcriptome sequencing, the group has so far published four articles in addition to papers III and IV using transcriptome sequencing to identify fusion genes (Panagopoulos *et al*, 2012, Panagopoulos *et al*, 2013a, Panagopoulos *et al*, 2013b, Panagopoulos *et al*, 2013c). Even though these studies for the most part were of single cases, we expect to broaden them in the future. To generate more knowledge and understanding, gene fusions should be examined in larger cohorts of patients with malignant diseases, preferably with the same cytogenetic characteristics, to determine their prevalence and specificity as biomarkers. For *HMGA2*, the gene we found to be deregulated in AML in paper II, the expression could be measured in a panel of AML patients with rearrangements of chromosome region 12q1.

The oncogenic potential of the involved genes can be investigated by functional studies as outlined in Table 4. Research in our laboratory is currently focused on describing the consequences of altered chromosomes rather than doing functional

analysis, so by publishing our results we hope that other researchers with more cell biological expertise will take the findings further.

Today, fusion gene detection in the clinical diagnostic setting is based on FISH (mostly interphase FISH) or RT-PCR. Either as an array, or using targeted next generation sequencing, it will soon be feasible to detect all known gene fusions in cancer in a single experiment, also applicable in a diagnostic setting (Lovf *et al*, 2011, Lovf *et al*, 2013). As outlined in the discussion, all variants of disease specific rearrangements are important to investigate, as subtypes of them may impact prognosis. An updated database including all these chimeric transcripts is necessary for such an approach, and much effort has already been made to catalogue chimeric transcripts in databases (Frenkel-Morgenstern *et al*, 2013, Mitelman *et al*, 2013).

In hematologic and lymphatic malignancies, monitoring of minimal residual disease can be performed by PCR analysis amplifying fusion genes (van der Velden *et al*, 2003). This technique has been successfully used for detection of *BCR-ABL* fusion gene transcripts in CML (Bagg, 2002). Recent studies have demonstrated that this method is applicable also in solid tumors. By next generation sequencing, Leary and co-workers identified tumor-specific rearrangements in breast and colorectal cancer patients and were able to assess disease status and drug responsiveness by monitoring these disease-specific gene rearrangements (Leary *et al*, 2010). Although still too expensive in most diagnostic settings, this approach demonstrates another potential benefit of studying acquired mutations in cancer, independently of the oncogenic properties they may possess.

A novel fusion gene such as *IRF2BP2-CDX1* (paper III) would probably be a good individual biomarker for the patient in whom it was identified, regardless of the limited knowledge of its function. A nascent relapse would most probable be detectable by PCR targeting this fusion gene. If such fusion gene detection could be implemented in a clinical setting, it would probably save both recourses associated with conducting and interpreting the evaluations and the adverse effects linked to recurrent radiological evaluations. However, disease monitoring using “individual”

fusion genes requires that the mutations in question are in fact "primary" or "drivers" and that they are present in both the primary tumor and at relapse. A recent paper reported that among genes found to be mutated, only ~30% were found throughout the tumor parenchyma (Gerlinger *et al*, 2012). One approach to meet this issue is to investigate a panel of mutated genes, where one or more most probably will be present in all tumor cells.

In order to obtain a better pathogenetic understanding of a specific tumor entity, a deeper investigation of one or a few well-characterized tumors using several methods (cytogenetics, RNA sequencing, gene expression profiling etc.) may reveal more about the pathogenesis than using one single approach applied to many, probably molecular distinct tumours (Liu, 2013). We hope our present and future studies of malignancies at different resolution levels will contribute to this development.

A final note to the therapeutic potential of chimeric proteins caused by chromosomal rearrangements will round off this work. First, a protein product has to be important for tumor cell survival in order to make a good target for treatment. Second, it is easier to target a protein that results in a gain-of-function and has enzymatic activity, than a protein with numerous weak interactions (Vogelstein *et al*, 2013). Although development of treatments directed against fusion genes may seem like a challenging task, we must be patient in our search. Four decades passed from the identification of the Philadelphia chromosome to imatinib was approved for therapy against CML, but it was definitely worth the wait (Druker, 2008).

List of abbreviations

ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
BAC	bacterial artificial chromosome
cDNA	complementary DNA
CGH	comparative genomic hybridization
CML	chronic myeloid leukemia
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
ESTs	expressed sequence tags
FAB	French American British
FISH	fluorescence *in situ* hybridization
GISTs	gastrointestinal stromal tumors
HLA	human leukocyte antigen
IP-FISH	interphase FISH
ISCN	International System for Human Cytogenetic Nomenclature
LD-PCR	long distance PCR
MC	mesenchymal chondrosarcoma
M-FISH	multiplex FISH
miRNA	micro RNA
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
PCR	polymerase chain reaction
qPCR	quantitative PCR (Real-time PCR)
RACE-PCR	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNA-seq	RNA sequencing
rRNA	ribosomal RNA
RT-PCR	reverse transcription
siRNA	small interfering
TKI	tyrosine kinase inhibitor
tRNA	transfer RNA
UCSC	University of California, Santa Cruz
WHO	World Health Organization

Gene symbols

ABL1	c-abl oncogene 1, non-receptor tyrosine kinase
ALK	anaplastic lymphoma receptor tyrosine kinase
BCR	breakpoint cluster region
CBFA2T3	core-binding factor, runt domain, alpha subunit 2; translocated to, 3
CDX1	caudal type homeobox 1
CDX2	caudal type homeobox 2
DDIT3	DNA-damage-inducible transcript 3
ETV6	ets variant 6
EWSR1	EWS RNA-binding protein 1
FLI1	Fli-1 proto-oncogene, ETS transcription factor
FLT3	fms-related tyrosine kinase 3
FUS	fused in sarcoma
GPC5	glypican 5
GUSB	glucuronidase, beta
HEY1	hairy/enhancer-of-split related with YRPW motif 1
HMGA2	high mobility group AT-hook 2
HMGA2	high mobility group AT-hook 1
IRF2	interferon regulatory factor 2
IRF2BP2	interferon regulatory factor 2 binding protein 2
JAK2	Janus kinase 2
KANK1	KN motif and ankyrin repeat domains 1
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
L1TD1	LINE-1 type transposase domain containing 1
MDM2	MDM2 oncogene, E3 ubiquitin protein ligase
MYC	v-myc avian myelocytomatosis viral oncogene homolog
NCOA2	nuclear receptor coactivator 2
NFIA	nuclear factor I/A
NTRK1	neurotrophic tyrosine kinase, receptor, type 1
PDGFRβ	platelet-derived growth factor receptor, beta polypeptide
PIGA	phosphatidylinositol glycan anchor biosynthesis, class A
RUNX1	runt-related transcription factor 1
TAF15	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa
TBP	TATA box binding protein
TNIK	TRAF2 and NCK interacting kinase
ZNF384	zinc finger protein 384

References

Aliano S, Cirmena G, Garuti A, Fugazza G, Bruzzone R, Rocco I *et al* (2007) HMGA2 overexpression in polycythemia vera with t(12;21)(q14;q22). *Cancer Genetics Cytogenetics*, 177(2):115-119.

American Cancer Society (2009) The history of cancer. American Cancer Society.

Andrieux J, Demory JL, Dupriez B, Quief S, Plantier I, Roumier C *et al* (2004) Dysregulation and overexpression of HMGA2 in myelofibrosis with myeloid metaplasia. *Genes Chromosomes Cancer*, 39(1):82-87.

Baccarani M, Saglio G, Goldman J, Hochhaus A, Simonsson B, Appelbaum F *et al* (2006) Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood*, 108(6):1809-1820.

Bagg A (2002) Chronic myeloid leukemia: a minimalistic view of post-therapeutic monitoring. *J Mol Diagn*, 4(1):1-10.

Balmain A (2001) Cancer genetics: from Boveri and Mendel to microarrays. *Nat Rev Cancer*, 1(1):77-82.

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A*, 91(6):2216-2220.

Bartlett JM, Stirling D (2003) A short history of the polymerase chain reaction. *Methods Mol Biol*, 226:3-6.

Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K *et al* (2011) Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat Genet*, 43(10):964-968.

Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E *et al* (2003) Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. *Leukemia*, 17(12):2474-2486.

Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR *et al* (1976) Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol*, 33(4):451-458.

Bjerkehagen B, Smeland S, Walberg L, Skjeldal S, Hall KS, Nesland JM *et al* (2008) Radiation-induced sarcoma: 25-year experience from the Norwegian Radium Hospital. *Acta Oncol*, 47(8):1475-1482.

Borden EC, Baker LH, Bell RS, Bramwell V, Demetri GD, Eisenberg BL *et al* (2003) Soft tissue sarcomas of adults: state of the translational science. *Clin Cancer Res*, 9(6):1941-1956.

Bose S, Deininger M, Gora-Tybor J, Goldman JM, Melo JV (1998) The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease. *Blood*, 92(9):3362-3367.

Braiteh F, Kurzrock R (2007) Uncommon tumors and exceptional therapies: paradox or paradigm? *Mol Cancer Ther*, 6(4):1175-1179.

Brandal P, Bjerkehagen B, Heim S (2004) Molecular cytogenetic characterization of tenosynovial giant cell tumors. *Neoplasia*, 6(5):578-583.

Brandal P, Teixeira MR, Heim S (2010) Genotypic and phenotypic classification of cancer: How should the impact of the two diagnostic approaches best be balanced? *Genes Chromosomes Cancer*, 49(9):763-774.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M *et al* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*, 55(4):611-622.

Calin GA, Croce CM (2007) Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J Clin Invest*, 117(8):2059-2066.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S *et al* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*, 101(9):2999-3004.

Cancer Registry of Norway (2013) Fakta om kreft [http://www.kreftregisteret.no/no/Generelt/Fakta-om-kreft-test/]

Carrara M, Beccuti M, Lazzarato F, Cavallo F, Cordero F, Donatelli S *et al* (2013) State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res Int*, 2013:340620.

Caspersson T, Farber S, Foley GE, Kudynowski J, Modest EJ, Simonsson E *et al* (1968) Chemical differentiation along metaphase chromosomes. *Exp Cell Res*, 49(1):219-222.

Castaneda VL, Parmley RT, Saldivar VA, Cheah MS (1991) Childhood undifferentiated leukemia with early erythroid markers and c-myb duplication. *Leukemia*, 5(2):142-149.

Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem*, 76:51-74.

Chase A, Reiter A, Burci L, Cazzaniga G, Biondi A, Pickard J *et al* (1999) Fusion of ETV6 to the caudal-related homeobox gene CDX2 in acute myeloid leukemia with the t(12;13)(p13;q12). *Blood*, 93(3):1025-1031.

Czader M, Orazi A (2009) Therapy-related myeloid neoplasms. *Am J Clin Pathol*, 132(3):410-425.

de Alava E, Kawai A, Healey JH, Fligman I, Meyers PA, Huvos AG *et al* (1998) EWS-FL11 fusion transcript structure is an independent determinant of prognosis in Ewing's sarcoma. *J Clin Oncology*, 16(4):1248-1255.

de Braekeleer E, Auffret R, Garcia JR, Padilla JM, Fletes CC, Morel F *et al* (2013) Identification of NIPBL, a new ETV6 partner gene in t(5;12) (p13;p13)-associated acute megakaryoblastic leukemia. *Leuk Lymphoma*, 54(2):423-424.

De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Basinko A, De Braekeleer M (2012) ETV6 fusion genes in hematological malignancies: a review. *Leuk Res*, 36(8):945-961.

de Klein A, van Kessel AG, Grosveld G, Bartram CR, Hagemeijer A, Bootsma D *et al* (1982) A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature*, 300(5894):765-767.

de la Grange P, Dutertre M, Correa M, Auboeuf D (2007) A new advance in alternative splicing databases: from catalogue to detailed analysis of regulation of expression and function of human alternative splicing variants. *BMC Bioinformatics*, 8:180.

Deininger MW, Goldman JM, Melo JV (2000) The molecular biology of chronic myeloid leukemia. *Blood*, 96(10):3343-3356.

Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M *et al* (1992) Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*, 359(6391):162-165.

Desmaze C, Brizard F, Turc-Carel C, Melot T, Delattre O, Thomas G *et al* (1997) Multiple chromosomal mechanisms generate an EWS/FLI1 or an EWS/ERG fusion gene in Ewing tumors. *Cancer Genet Cytogenet*, 97(1):12-19.

Domingo-Claros A, Larriba I, Rozman M, Irriguible D, Vallespi T, Aventin A *et al* (2002) Acute erythroid neoplastic proliferations. A biological study based on 62 patients. *Haematologica*, 87(2):148-153.

Druker BJ (2002) Inhibition of the Bcr-Abl tyrosine kinase as a therapeutic strategy for CML. *Oncogene*, 21(56):8541-8546.

Druker BJ (2008) Translation of the Philadelphia chromosome into therapy for CML. *Blood*, 112(13):4808-4817.

Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S *et al* (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med*, 2(5):561-566.

Edwards PA (2010) Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol*, 220(2):244-254.

Eguchi M, Eguchi-Ishimae M, Tojo A, Morishita K, Suzuki K, Sato Y *et al* (1999) Fusion of ETV6 to neurotrophin-3 receptor TRKC in acute myeloid leukemia with t(12;15)(p13;q25). *Blood*, 93(4):1355-1363.

Epstein L, DeVries S, Waldman FM (1995) Reutilization of previously hybridized slides for fluorescence in situ hybridization. *Cytometry*, 21(4):378-381.

Faderl S, Kantarjian HM, Talpaz M, Estrov Z (1998) Clinical significance of cytogenetic abnormalities in adult acute lymphoblastic leukemia. *Blood*, 91(11):3995-4019.

Fan Y-S (2002) Molecular Cytogenetics; Protocols and Applications, vol. 204. Totowa, New Jersey: Humana Press.

Ferlin A, Raicu F, Gatta V, Zuccarello D, Palka G, Foresta C (2007) Male infertility: role of genetic background. *Reprod Biomed Online*, 14(6):734-745.

Ferner RE, Gutmann DH (2002) International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis. *Cancer Res*, 62(5):1573-1577.

Fletcher CDM, J.A. B, Hogendoorn P.C.W, Mertens F (eds.): WHO Classification of Tumours of Soft Tissue and Bone. Lyon: IARCPress; 2013.

Frenkel-Morgenstern M, Gorohovski A, Lacroix V, Rogers M, Ibanez K, Boullosa C *et al* (2013) ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. *Nucleic Acids Res*, 41(Database issue):D142-151.

Furney SJ, Higgins DG, Ouzounis CA, Lopez-Bigas N (2006) Structural and functional properties of genes involved in human cancer. *BMC Genomics*, 7:3.

Fusco A, Fedele M (2007) Roles of HMGA proteins in cancer. *Nat Rev Cancer*, 7(12):899-910.

Galbraith D, Gross SA, Paustenbach D (2010) Benzene and human health: A historical review and appraisal of associations with various diseases. *Crit Rev Toxicol*, 40 Suppl 2:1-46.

Gamou T, Kitamura E, Hosoda F, Shimizu K, Shinohara K, Hayashi Y *et al* (1998) The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood*, 91(11):4028-4037.

Gatter KM, Olson S, Lawce H, Rader AE (2005) Trisomy 8 as the sole cytogenetic abnormality in a case of extraskeletal mesenchymal chondrosarcoma. *Cancer Genet Cytogenet*, 159(2):151-154.

Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W (2011) FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics*, 27(14):1922-1928.

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E *et al* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, 366(10):883-892.

Giacomini CP, Sun S, Varma S, Shain AH, Giacomini MM, Balagtas J *et al* (2013) Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. *PLoS Genet*, 9(4):e1003464.

Gianfelici V, Lahortiga I, Cools J (2012) Chromosomal aberrations and fusion genes in myeloid malignancies. *Expert Rev Hematol*, 5(4):381-393.

Gibson UE, Heid CA, Williams PM (1996) A novel method for real time quantitative RT-PCR. *Genome Res*, 6(10):995-1001.

Gilliland DG, Griffin JD (2002) The roles of FLT3 in hematopoiesis and leukemia. *Blood*, 100(5):1532-1542.

Goldman JM (2010) Chronic myeloid leukemia: a historical perspective. *Semin Hematol*, 47(4):302-311.

Gonzalez I, Vicent S, de Alava E, Lecanda F (2007) EWS/FLI-1 oncoprotein subtypes impose different requirements for transformation and metastatic activity in a murine model. *J Mol Med (Berl)*, 85(9):1015-1029.

Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH *et al* (2010) Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*, 116(3):354-365.

Groffen J, Stephenson JR, Heisterkamp N, de Klein A, Bartram CR, Grosveld G (1984) Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell*, 36(1):93-99.

Group Francais de Cytogenetique Hematologique (1996) Cytogenetic abnormalities in adult acute lymphoblastic leukemia: correlations with hematologic findings outcome. A Collaborative Study of the Group Francais de Cytogenetique Hematologique, vol. 87, 1996/04/15 edn.

Harrison CJ (2009) Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia. *Br J Haematol*, 144(2):147-156.

Harrison CJ (2011) Acute lymphoblastic leukemia. *Clin Lab Med*, 31(4):631-647, ix.

Harrison CJ, Moorman AV, Barber KE, Broadfield ZJ, Cheung KL, Harris RL *et al* (2005) Interphase molecular cytogenetic screening for chromosomal abnormalities of prognostic significance in childhood acute lymphoblastic leukaemia: a UK Cancer Cytogenetics Group Study. *Br J Haematol*, 129(4):520-530.

Harrison CJ, Moorman AV, Broadfield ZJ, Cheung KL, Harris RL, Reza Jalali G *et al* (2004) Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *Br J Haematol*, 125(5):552-559.

Hecht JL, Aster JC (2000) Molecular biology of Burkitt's lymphoma. *J Clin Oncol*, 18(21):3707-3721.

Heim S, Mitelman F (1989) Primary chromosome abnormalities in human neoplasia. *Adv Cancer Res*, 52:1-43.

Heim S, Mitelman F (2009) Cancer cytogenetics, 3rd ed. edn. Hoboken, N.J. :: Wiley-Blackwell.

Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen CJ, Joseph N *et al* (2003) PDGFRA activating mutations in gastrointestinal stromal tumors. *Science*, 299(5607):708-710.

Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S *et al* (1998) Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science*, 279(5350):577-580.

Howarth KD, Pole JC, Beavis JC, Batty EM, Newman S, Bignell GR *et al* (2011) Large duplications at reciprocal translocation breakpoints that might be the counterpart of large deletions and could arise from stalled replication bubbles. *Genome Res*, 21(4):525-534.

Inoue N, Izui-Sarumaru T, Murakami Y, Endo Y, Nishimura J, Kurokawa K *et al* (2006) Molecular basis of clonal expansion of hematopoiesis in 2 patients with paroxysmal nocturnal hemoglobinuria (PNH). *Blood*, 108(13):4232-4236.

Joensuu H, Roberts PJ, Sarlomo-Rikala M, Andersson LC, Tervahartiala P, Tuveson D *et al* (2001) Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med*, 344(14):1052-1056.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F *et al* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 258(5083):818-821.

Kampen KR (2012) The discovery and early understanding of leukemia. *Leuk Res*, 36(1):6-13.

Kapur K, Xing Y, Ouyang Z, Wong WH (2007) Exon arrays provide accurate assessments of gene expression. *Genome Biol*, 8(5):R82.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM *et al* (2002) The human genome browser at UCSC. *Genome Res*, 12(6):996-1006.

Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res*, 20(5):1083-1085.

Knezevich SR, McFadden DE, Tao W, Lim JF, Sorensen PH (1998) A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nat Genet*, 18(2):184-187.

Kochetkova M, McKenzie OL, Bais AJ, Martin JM, Secker GA, Seshadri R *et al* (2002) CBFA2T3 (MTG16) is a putative breast tumor suppressor gene from the breast cancer loss of heterozygosity region at 16q24.3. *Cancer Res*, 62(16):4599-4604.

Koller U, Haas OA, Ludwig WD, Bartram CR, Harbott J, Panzer-Grumayer R *et al* (1989) Phenotypic and genotypic heterogeneity in infant acute leukemia. II. Acute nonlymphoblastic leukemia. *Leukemia*, 3(10):708-714.

Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA *et al* (2001) The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood*, 98(6):1752-1759.

Kottickal LV, Sarada B, Ashar H, Chada K, Nagarajan L (1998) Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal in human leukemia. *Biochem Biophys Res Commun*, 242(2):452-456.

Kutok JL, Aster JC (2002) Molecular biology of anaplastic lymphoma kinase-positive anaplastic large-cell lymphoma. *J Clin Oncol*, 20(17):3691-3702.

Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG *et al* (2010) Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*, 363(18):1693-1703.

La Starza R, Aventin A, Crescenzi B, Gorello P, Specchia G, Cuneo A *et al* (2005) CIZ gene rearrangements in acute leukemia: report of a diagnostic FISH assay and clinical features of nine patients. *Leukemia*, 19(9):1696-1699.

Landegent JE, Jansen in de Wal N, van Ommen GJ, Baas F, de Vijlder JJ, van Duijn P *et al* (1985) Chromosomal localization of a unique gene by non-autoradiographic in situ hybridization. *Nature*, 317(6033):175-177.

Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C *et al* (2010) Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med*, 2(20):20ra14.

Lee C, Gisselsson D, Jin C, Nordgren A, Ferguson DO, Blennow E *et al* (2001) Limitations of chromosome classification by multicolor karyotyping. *Am J Hum Genet*, 68(4):1043-1047.

Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K *et al* (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, 456(7218):66-72.

Li FP, Fraumeni JF, Jr., Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA *et al* (1988) A cancer family syndrome in twenty-four kindreds. *Cancer Res*, 48(18):5358-5362.

Liu ET (2013) Grappling with cancer. *Science*, 339(6127):1493.

Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl*, 4(6):357-362.

Lovf M, Thomassen GO, Bakken AC, Celestino R, Fioretos T, Lind GE *et al* (2011) Fusion gene microarray reveals cancer type-specificity among fusion genes. *Genes Chromosomes Cancer*, 50(5):348-357.

Lovf M, Thomassen GO, Mertens F, Cerveira N, Teixeira MR, Lothe RA *et al* (2013) Assessment of fusion gene status in sarcomas using a custom made fusion gene microarray. *PLoS One*, 8(8):e70649.

Lugo TG, Pendergast AM, Muller AJ, Witte ON (1990) Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science*, 247(4946):1079-1082.

Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X *et al* (2009) Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458(7234):97-101.

Martini A, La Starza R, Janssen H, Bilhou-Nabera C, Corveleyn A, Somers R *et al* (2002) Recurrent rearrangement of the Ewing's sarcoma gene, EWSR1, or its homologue, TAF15, with the transcription factor CIZ/NMP4 in acute leukemia. *Cancer Res*, 62(19):5408-5412.

Mayr C, Hemann MT, Bartel DP (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science*, 315(5818):1576-1579.

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG *et al* (2011) deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol*, 7(5):e1001138.

Meyer B, Krisponeit D, Junghanss C, Murua Escobar H, Bullerdiek J (2007) Quantitative expression analysis in peripheral blood of patients with chronic myeloid leukaemia: correlation between HMGA2 expression and white blood cell count. *Leuk Lymphoma*, 48(10):2008-2013.

Micci F, Panagopoulos I, Bjerkehagen B, Heim S (2006) Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. *Cancer Res*, 66(1):107-112.

Micci F, Teixeira MR, Bjerkehagen B, Heim S (2002) Characterization of supernumerary rings and giant marker chromosomes in well-differentiated lipomatous tumors by a combination of G-banding, CGH, M-FISH, and chromosome- and locus-specific FISH. *Cytogenet Genome Res*, 97(1-2):13-19.

Micci F, Thorsen J, Haugom L, Zeller B, Tierens A, Heim S (2011) Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. *Leukemia*, 25(9):1510-1512.

Mitelman F (2000) Recurrent chromosome aberrations in cancer. *Mutat Res*, 462(2-3):247-253.

Mitelman F, Johansson B, Mertens F (2004) Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat Genet*, 36(4):331-334.

Mitelman F, Johansson B, Mertens F (2007) The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer*, 7(4):233-245.

Mitelman F, Johansson B, Mertens F, (Eds) (2013) Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [http://cgap.nci.nih.gov/Chromosomes/Mitelman]

Moorman AV (2012) The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Rev*, 26(3):123-135.

Moorman AV, Harrison CJ, Buck GA, Richards SM, Secker-Walker LM, Martineau M *et al* (2007) Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood*, 109(8):3189-3197.

Moyzis RK, Albright KL, Bartholdi MF, Cram LS, Deaven LL, Hildebrand CE *et al* (1987) Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis. *Chromosoma*, 95(6):375-386.

Mukherjee S (2011) The Emperor of All Maladies: Fourth Estate.

Munoz J, Kurzrock R (2012) Targeted therapy in rare cancers--adopting the orphans. *Nat Rev Clin Oncol*, 9(11):631-642.

Nakashima Y, Unni KK, Shives TC, Swee RG, Dahlin DC (1986) Mesenchymal chondrosarcoma of bone and soft tissue. A review of 111 cases. *Cancer*, 57(12):2444-2453.

Nakayama R, Miura Y, Ogino J, Susa M, Watanabe I, Horiuchi K *et al* (2012) Detection of HEY1-NCOA2 fusion by fluorescence in-situ hybridization in formalin-fixed paraffin-embedded tissues as a possible diagnostic tool for mesenchymal chondrosarcoma. *Pathol Int*, 62(12):823-826.

Naumann S, Krallman PA, Unni KK, Fidler ME, Neff JR, Bridge JA (2002) Translocation der(13;21)(q10;q10) in skeletal and extraskeletal mesenchymal chondrosarcoma. *Mod Pathol*, 15(5):572-576.

Nowell PC, Hungerford DA (1960a) Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst*, 25:85-109.

Nowell PC, Hungerford DA (1960b) Minute chromosome in human chronic granulocytic leukemia. *Science*, 132(3438):1497-1497.

Odero MD, Grand FH, Iqbal S, Ross F, Roman JP, Vizmanos JL *et al* (2005) Disruption and aberrant expression of HMGA2 as a consequence of diverse chromosomal translocations in myeloid malignancies. *Leukemia*, 19(2):245-252.

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ *et al* (2001) A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res*, 11(3):483-496.

Panagopoulos I, Micci F, Thorsen J, Haugom L, Buechner J, Kerndrup G *et al* (2013a) Fusion of ZMYND8 and RELA Genes in Acute Erythroid Leukemia. *PLoS One*, 8(5):e63663.

Panagopoulos I, Micci F, Thorsen J, Haugom L, Tierens A, Ulvmoen A *et al* (2012) A novel TCF3-HLF fusion transcript in acute lymphoblastic leukemia with a t(17;19)(q22;p13). *Cancer Genet*, 205(12):669-672.

Panagopoulos I, Thorsen J, Gorunova L, Haugom L, Bjerkehagen B, Davidson B *et al* (2013b) Fusion of the ZC3H7B and BCOR genes in endometrial stromal sarcomas carrying an X;22-translocation. *Genes Chromosomes Cancer*, 52(7):610-618.

Panagopoulos I, Thorsen J, Gorunova L, Micci F, Haugom L, Davidson B *et al* (2013c) RNA sequencing identifies fusion of the EWSR1 and YY1 genes in mesothelioma with t(14;22)(q32;q12). *Genes Chromosomes Cancer*, 52(8):733-740.

Pardue ML, Gall JG (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc Natl Acad Sci U S A*, 64(2):600-604.

Perez-Losada J, Pintado B, Gutierrez-Adan A, Flores T, Banares-Gonzalez B, del Campo JC *et al* (2000) The chimeric FUS/TLS-CHOP fusion protein specifically induces liposarcomas in transgenic mice. *Oncogene*, 19(20):2413-2422.

Pierantoni GM, Santulli B, Caliendo I, Pentimalli F, Chiappetta G, Zanesi N *et al* (2003) HMGA2 locus rearrangement in a case of acute lymphoblastic leukemia. *Int J Oncol*, 23(2):363-367.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D *et al* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*, 20(2):207-211.

Pui CH, Evans WE (2006) Treatment of acute lymphoblastic leukemia. *N Engl J Med*, 354(2):166-178.

Raimondi SC (2000) Fluorescence in situ hybridization: molecular probes for diagnosis of pediatric neoplastic diseases. *Cancer Invest*, 18(2):135-147.

Richkind KE, Romansky SG, Finklestein JZ (1996) t(4;19)(q35;q13.1): a recurrent change in primitive mesenchymal tumors? *Cancer Genet Cytogenet*, 87(1):71-74.

Rigby PW, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J Mol Biol*, 113(1):237-251.

Riggi N, Cironi L, Provero P, Suva ML, Kaloulis K, Garcia-Echeverria C *et al* (2005) Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. *Cancer Res*, 65(24):11459-11468.

Rommel B, Rogalla P, Jox A, Kalle CV, Kazmierczak B, Wolf J *et al* (1997) HMGI-C, a member of the high mobility group family of proteins, is expressed in hematopoietic stem cells and in leukemic cells. *Leuk Lymphoma*, 26(5-6):603-607.

Rowley JD (1973) Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243(5405):290-293.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA *et al* (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230(4732):1350-1354.

Sambrook J, Russel DW 2001 In Vitro Amplification of DNA by the polymerase Chain Reaction. In: *Molecular cloning: a laboratory manual.* 3. edn: Cold Spring Harbor Laboratory Press.

Sambrook J, Russell D 2001 Protocol 9 Rapid Amplification of 5`cDNA ends. In: *Molecular Cloning: A Laboratory Manual.* Edited by J S, DW R. Cold Spring Harbor, New York, USA: Cold Spring Harbor Laboratory Press: 8.54-58.60.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA *et al* (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265(5596):687-695.

Santulli B, Kazmierczak B, Napolitano R, Caliendo I, Chiappetta G, Rippe V *et al* (2000) A 12q13 translocation involving the HMGI-C gene in richter transformation of a chronic lymphocytic leukemia. *Cancer Genet Cytogenet*, 119(1):70-73.

Sawyers C (2004) Targeted cancer therapy. *Nature*, 432(7015):294-297.

Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA *et al* (1996) Multicolor spectral karyotyping of human chromosomes. *Science*, 273(5274):494-497.

Scotto-Lavino E, Du G, Frohman MA (2006a) 3' end cDNA amplification using classic RACE. *Nat Protoc*, 1(6):2742-2745.

Scotto-Lavino E, Du G, Frohman MA (2006b) Amplification of 5' end cDNA with 'new RACE'. *Nat Protoc*, 1(6):3056-3061.

Shaffer LG, Slovak ML, Campbell LJ, Nomenclature ISCoHC (2009) ISCN 2009: An International System for Human Cytogenetic Nomenclature (2009): Karger Medical Scientific.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol*, 26(10):1135-1145.

Shendure J, Lieberman Aiden E (2012) The expanding scope of DNA sequencing. *Nat Biotechnol*, 30(11):1084-1094.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y *et al* (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A*, 89(18):8794-8797.

Sierra JR, Cepero V, Giordano S (2010) Molecular mechanisms of acquired resistance to tyrosine kinase targeted therapy. *Mol Cancer*, 9:75.

Sinclair PB, Nacheva EP, Leversha M, Telford N, Chang J, Reid A *et al* (2000) Large deletions at the t(9;22) breakpoint are common and may identify a poor-prognosis subgroup of patients with chronic myeloid leukemia. *Blood*, 95(3):738-743.

Smeets DF (2004) Historical prospective of human cytogenetics: from microscope to microarray. *Clin Biochem*, 37(6):439-446.

Smith A, Howell D, Patmore R, Jack A, Roman E (2011) Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. *Br J Cancer*, 105(11):1684-1692.

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S *et al* (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*, 448(7153):561-566.

Speicher MR, Carter NP (2005) The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet*, 6(10):782-792.

Speicher MR, Gwyn Ballard S, Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet*, 12(4):368-375.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT *et al* (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*, 462(7276):1005-1010.

Storlazzi CT, Albano F, Locunsolo C, Lonoce A, Funes S, Guastadisegni MC *et al* (2006) t(3;12)(q26;q14) in polycythemia vera is associated with upregulation of the HMGA2 gene. *Leukemia*, 20(12):2190-2192.

Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature*, 458(7239):719-724.

Swerdlow S, Campo E, Harris NL (2001) WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press.

Swerdlow SH, Campo, E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Vardiman, J.W (ed.): WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, 4 edn. Lyon: IARC 2008.

Tan AY, Manley JL (2009) The TET family of proteins: functions and roles in disease. *J Mol Cell Biol*, 1(2):82-92.

Tanas MR, Sboner A, Oliveira AM, Erickson-Johnson MR, Hespelt J, Hanwright PJ *et al* (2011) Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma. *Sci Transl Med*, 3(98):98ra82.

Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M (2011) Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer*, 11(8):541-557.

Teixeira MR (2002) Combined classical and molecular cytogenetic analysis of cancer. *Eur J Cancer*, 38(12):1580-1584.

Teixeira MR, Micci F, Dietrich CU, Heim S (2000) Detailed genome-wide screening for inter- and intrachromosomal abnormalities by sequential G-banding and RxFISH color banding of the same metaphase cells. *Cancer Genet Cytogenet*, 119(2):94-101.

Tjio JH, Levan A (1956) The chromosome number of man. *Hereditas*, 42:1–6.

Tognon C, Knezevich SR, Huntsman D, Roskelley CD, Melnyk N, Mathers JA *et al* (2002) Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell*, 2(5):367-376.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW *et al* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*, 310(5748):644-648.

Trask BJ (2002) Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet*, 3(10):769-778.

van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJ (2003) Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia*, 17(6):1013-1034.

Vandenbroucke JP (2001) In defense of case reports and case series. *Ann Intern Med*, 134(4):330-334.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*, 10(4):252-263.

Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A *et al* (2009) The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*, 114(5):937-951.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW (2013) Cancer genome landscapes. *Science*, 339(6127):1546-1558.

Wang L, Motoi T, Khanin R, Olshen A, Mertens F, Bridge J *et al* (2012) Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data. *Genes Chromosomes Cancer*, 51(2):127-139.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1):57-63.

Werner M, Wilkens L, Aubele M, Nolte M, Zitzelsberger H, Komminoth P (1997) Interphase cytogenetics in pathology: principles, methods, and applications of fluorescence in situ hybridization (FISH). *Histochem Cell Biol*, 108(4-5):381-390.

Williamson D, Selfe J, Gordon T, Lu YJ, Pritchard-Jones K, Murai K *et al* (2007) Role for amplification and expression of glypican-5 in rhabdomyosarcoma. *Cancer Res*, 67(1):57-65.

Zeeb H, Blettner M (1998) Adult leukaemia: what is the role of currently known risk factors? *Radiat Environ Biophys*, 36(4):217-228.

Zoubek A, Dockhorn-Dworniczak B, Delattre O, Christiansen H, Niggli F, Gatterer-Menz I *et al* (1996) Does expression of different EWS chimeric transcripts define clinically distinct risk groups of Ewing tumor patients? *J Clin Oncol*, 14(4):1245-1251.

Paper I

Kaja B. Nyquist, Jim Thorsen, Bernward Zeller, Alf Haaland, Gunhild Trøen, Sverre Heim, Francesca Micci:

Identification of the *TAF15-ZNF384* fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12).

Cancer Genetics (2011) 204(3):147-152.

Paper II

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Roberta Roberto, Hilde S. Wik, Anne Tierens, Sverre Heim, Francesca Micci:

t(12;13)(q14;q31) leading to *HMGA2* upregulation in acute myeloid leukaemia.

Correspondence to British Journal of Haematology (2012), 157(6):769–771.

Paper III

Kaja B. Nyquist, Ioannis Panagopoulos, Jim Thorsen, Lisbeth Haugom, Ludmila Gorunova, Bodil Bjerkehagen, Alexander Fosså, Marianne Guriby, Torfinn Nome, Ragnhild A. Lothe, Rolf I. Skotheim, Sverre Heim, Francesca Micci:

Whole-transcriptome sequencing identifies novel *IRF2BP2-CDX1* fusion gene brought about by translocation t(1;5)(q42;q32) in mesenchymal chondrosarcoma.

PLOS ONE (2012), 7(11):e49705.

OPEN ACCESS Freely available online

PLOS ONE

Whole-Transcriptome Sequencing Identifies Novel *IRF2BP2-CDX1* Fusion Gene Brought about by Translocation t(1;5)(q42;q32) in Mesenchymal Chondrosarcoma

Kaja B. Nyquist[1,2,3]*, Ioannis Panagopoulos[1,2], Jim Thorsen[1,2], Lisbeth Haugom[1,2], Ludmila Gorunova[1,2], Bodil Bjerkehagen[4], Alexander Fossså[5], Marianne Guriby[2,6], Torfinn Nome[2,6], Ragnhild A. Lothe[2,6], Rolf I. Skotheim[2,6], Sverre Heim[1,2,3], Francesca Micci[1,2]

1 Section for Cancer Cytogenetics, Institute for Medical Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **2** Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway, **3** Faculty of Medicine, University of Oslo, Oslo, Norway, **4** Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **5** Department of Oncology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, **6** Department of Cancer Prevention, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

Abstract

Mesenchymal chondrosarcomas (MCs) account for 3–10% of primary chondrosarcomas. The cytogenetic literature includes only ten such tumours with karyotypic information and no specific aberrations have been identified. Using a purely molecular genetic approach a *HEY1-NCOA2* fusion gene was recently detected in 10 of 15 investigated MCs. The fusion probably arises through intrachromosomal rearrangement of chromosome arm 8 q. We report a new case of MC showing a t(1;5)(q42;q32) as the sole karyotypic aberration. Through FISH and whole transcriptome sequencing analysis we found a novel fusion between the *IRF2BP2* gene and the transcription factor *CDX1* gene arising from the translocation. The *IRF2BP2-CDX1* has not formerly been described in human neoplasia. In our hospital's archives three more cases of MC were found, and we examined them looking for the supposedly more common *HEY1-NCOA2* fusion, finding it in all three tumours but not in the case showing t(1;5) and *IRF2BP2-CDX1* gene fusion. This demonstrates that genetic heterogeneity exists in mesenchymal chondrosarcoma.

Citation: Nyquist KB, Panagopoulos I, Thorsen J, Haugom L, Gorunova L, et al. (2012) Whole-Transcriptome Sequencing Identifies Novel *IRF2BP2-CDX1* Fusion Gene Brought about by Translocation t(1;5)(q42;q32) in Mesenchymal Chondrosarcoma. PLoS ONE 7(11): e49705. doi:10.1371/journal.pone.0049705

Editor: Kwok-Wai Lo, The Chinese University of Hong Kong, Hong Kong

Received June 28, 2012; **Accepted** October 9, 2012; **Published** November 21, 2012

Funding: The study was supported by grants from the Norwegian Cancer Society and the South-Eastern Norway Regional Health Authority. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: k.b.nyquist@medisin.uio.no

Introduction

The classification of sarcomas describes over 50 different histological subtypes [1]. In approximately 20% of them, recurrent balanced translocations leading to formation of fusion genes were identified [2]. Fusion genes provide diagnostic and sometimes prognostic information on the tumours they characterize and knowledge about them could ultimately lead to new targeted therapies [3].

Mesenchymal chondrosarcomas (MCs) are rare tumours that account for 3–10% of primary chondrosarcomas [1]. Their typical histological appearance includes a biphasic pattern with areas of round primitive mesenchymal cells interrupted by chondroid elements [4]. Most cases are diagnosed in the second and third decade of life and the prognosis is mostly poor, with a 5-year survival rate of about 50% [5]. Adequate surgery is the gold standard for treatment of localized disease [6] and the role of chemotherapy and radiotherapy remains poorly defined [7,8].

According to the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [9], only ten MCs have been karyotyped and no consistent cytogenetic findings have been described. Recently, however, using a genome-wide exon-resolution expression screen, a fusion between the hairy/enhancer-of-split related with YRPW motif 1 *(HEY1)* gene and the nuclear receptor coactivator 2 *(NCOA2)* gene was detected in 10 out of 15 analysed MCs (67%) [10]. Both genes are located on the long arm of chromosome 8 and so the fusion presumably results from an intrachromosomal rearrangement, probably a deletion (~9.6 Mb according to the UCSC browser, assembly 2009).

We report the finding of a balanced t(1;5)(q42;q32) as the sole karyotypic abnormality in an MC. The translocation led to a new fusion between the interferon regulatory factor 2 binding protein 2 gene *(IRF2BP2)* and the caudal type homeobox 1 *(CDX1)* gene. Based on the recent report by Wang et al (2012) [10], we also examined archival material from another three MCs we had access to for the presence of the *HEY1-NCOA2* gene fusion, finding it in all three.

Materials and Methods

Patient Samples

Patient 1 was a 63-year-old female in whom a solitary tumour mass was detected in the right cerebral hemisphere in August 2007. Examination of biopsy material revealed the tumour to be a diffuse large B-cell lymphoma of activated B-cell subtype. Cytogenetic analysis of this tumour was unsuccessful. Detailed work-up for other manifestations of lymphoma was negative, compatible with a diagnosis of primary central nervous system lymphoma (PCNSL). However, a tumour in the left iliacus muscle was detected, 3 cm in largest diameter. Biopsies revealed a spindle cell tumour of uncertain malignant potential. The patient received chemotherapy for PCNSL according to Abrey et al. [11] including high-dose methotrexate and high-dose cytarabin. Evaluation after 7 courses of chemotherapy confirmed complete remission of her PCNSL. There was no change in size of the tumour in the left iliac muscle and in June 2008 a wide excision of it was performed. A detailed work up of the tumour specimen revealed a small cell and chondromatous tumour diagnosed as a mesenchymal chondrosarcoma (Figure 1A). Focal infiltrative growth and necroses were present. Because of narrow margins, postoperative radiotherapy 2 Gy × 25 was given. A CNS recurrence of her lymphoma was detected in November 2011, and the patient has received radiation therapy. She remains without sign of recurrence of the MC at the time of writing.

The Norwegian Radium Hospital (NRH) is the largest referral centre for Norwegian patients with bone and soft tissue tumours covering a population of 2.6 million. To identify additional patients with a diagnosis of MC, a database search was performed for cases with this disease. Three additional patients (patients 2–4) were identified (see Table 1 for clinical details and Figure 1A for histological image).

Ethics Statement

Written informed consent was obtained from patients 1 and 3. In the latter case, one of the parents consented on the patient's behalf. Frozen tissue from deceased patients (patients 2 and 4) was retrieved from The Radium Hospital biobank (project nr S-0747a approved by the Regional Ethics Committee and The Directory of Health (Helsedirektorat) in 2008). Specific permission to perform RNA analysis/sequencing was obtained from patient 1 after approval by the Regional Ethics Committee for Medical and Health Research Ethics South-East (REC number: 2010/1389A). The entire study was also approved by the institutional review board at the Norwegian Radium Hospital. All data were analyzed anonymously.

G-banding and karyotyping

Fresh tissue from a representative area of the tumour (patient 1) was received and analysed as part of our diagnostic routine. The samples were disaggregated mechanically and enzymatically with collagenase II (Worthington, Freehold, NJ, USA). The resulting cells were cultured and harvested using standard cytogenetic techniques [12]. Chromosome preparations were G-banded with Wright stain. The karyotype was written according to The International System for Human Cytogenetic Nomenclature (ISCN) 2009 guidelines [13]. Phytohemagglutinin (VWR, Oslo, Norway) –stimulated leucocytes were obtained from peripheral blood to determine the patient's constitutional karyotype.

Fluorescence in Situ Hybridization Analysis

Fluorescence in situ hybridization (FISH) was performed using probes from bacterial artificial chromosomes (BACs). BACs and fosmid clones flanking and covering the breakpoint positions were selected using the Human Genome Browser at the University of California web site (Feb.2009/release: hg19, http://genome.ucsc.edu/). The selected clones (see Table S1 for detailed information) were purchased from Life Technologies (Carlsbad, CA, USA) or the BACPAC Resource Center (Oakland, CA, USA).

Bacteria were cultured in selective media according to the manufacturer's recommendation. DNA was extracted using High Pure Plasmid Isolation kit (Roche Applied Science, Penzberg, Germany). DNA labelling was done in a nick translation reaction and the synthesized probes were hybridized to previously G-banded slides. All procedures were performed as previously described [14]. The slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). The analysis was done using a CytoVision system (Applied Imaging, Newcastle, UK). All probes were tested for their correct location on normal metaphase spreads prior to use.

Material for Molecular Analysis

Representative samples of tumour tissue were frozen and stored at −80°C after surgery. DNA was isolated using Genomic-tip (Qiagen, Hilden, Germany) to obtain pure high molecular weight DNA. RNA was extracted from tumour tissue using the Trizol reagent (Life Technologies) with a homogenizer (Omni THQ Digital Tissue Homogenizer, Kennesaw, GA, USA). The RNA quality was evaluated using the Experion Automated Electrophoresis System (Bio-Rad Laboratories, Hercules, CA, USA). cDNA was synthesized using the iScript kit and random primers (Bio-Rad Laboratories). All procedures were done according to the manufacturers' recommendations.

High-throughput Paired-end RNA-sequencing

Sequencing was performed according to the TruSeq paired-end RNA-sequencing protocols from Illumina for Solexa sequencing on a Genome Analyzer IIx with paired end module (Illumina Inc., San Diego, CA, USA). 3.5 μg total RNA was used as starting material for library construction, using the TruSeq RNA Sample Preparation Kit v2 where the steps include poly-A mRNA isolation, fragmentation, and cDNA synthesis before adapters are ligated to the products and amplified to a final cDNA library. Shearing to about 150 bp fragments was achieved using divalent cations under elevated temperature. Approximately 58 million clusters were generated by the TruSeq PE Cluster Kit v2 on the Illumina cBot Cluster Generation System, and 76 base pairs were sequenced, from each side of the fragments, using reagents from the TruSeq SBS Kit v5 (all kits from Illumina).

Gene Fusion Prediction

The Illumina software pipeline was used for processing of image data into raw sequencing data (SCS 2.9 and Casava 1.8.2), and only sequence reads marked as "passed filtering" were used in the downstream data analysis. A total of 91 million reads were obtained. We utilized the fusion discovery software deFuse (version 0.4.3) [15], with Ensembl release 65 reference genome (hg19) and gene models, RepeatMasker, EST, and spliced EST annotations downloaded from the University of California Santa Cruz Table Browser (http://genome.ucsc.edu/, accessed May 2012). UniGene clusters were downloaded from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/, accessed May 2012) to assist in locating potential gene fusions. Three spanning reads and two split reads were required to call sequence reads a gene fusion.

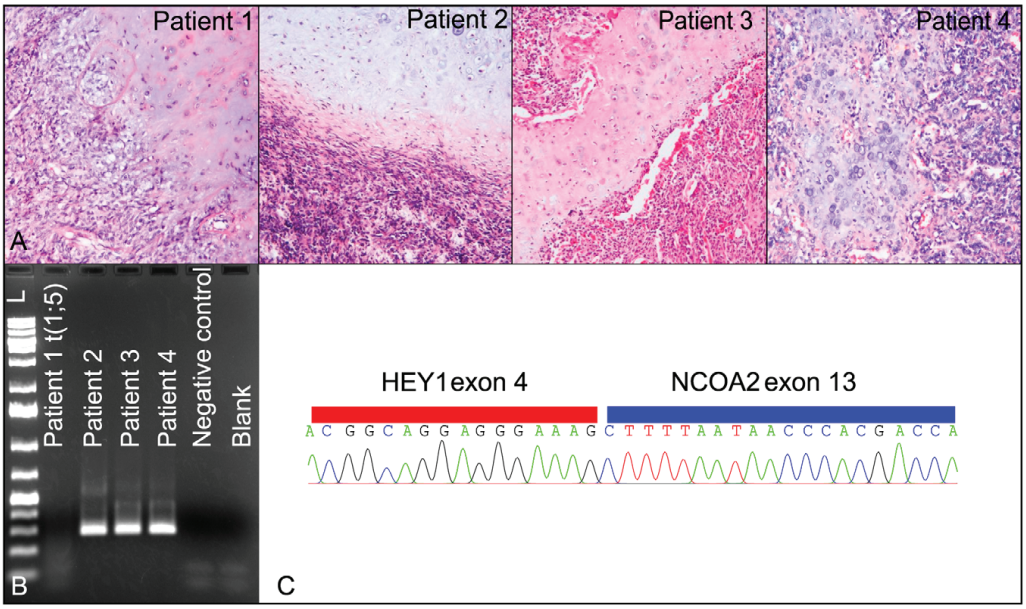


Figure 1. Histological images of the four MCs and characteristics of the *HEY1-NCOA2* fusion. (A) The typical biphasic histological pattern is observed in all tumours. (B) The *HEY1-NCOA2* fusion was detected using primers HEY1_F1 and NCOA2_E13-R3 in patients 2–4 but not in patient 1 whose tumour showed the t(1;5). (C) The *HEY1-NCOA2* fusion was confirmed by sequencing. The breakpoint positions were identical to those previously reported.
doi:10.1371/journal.pone.0049705.g001

PCR and Sequencing

Primers used in PCR were designed with the FastPCR software [16]. The full list of applied primers is given in Table 2. The primers used for detection of the *HEY1- NCOA2* fusion were identical to the primers used by Wang et al [10]. cDNA PCR was run using 2 µl cDNA in a 25 µl PCR reaction using TaKaRa Ex Taq Hot Start (Takara Bio Inc, Shiga, Japan). The PCR conditions were as follows: 98°C for 7 sec, 68°C for 2 min after a 1 min initial denaturation at 98°C. 34 cycles were run. Amplified products were cloned using the TOPO TA cloning kit (Life Technologies). Selected products were sent for Sanger sequencing (GATC Biotech, Konstanz, Germany) and obtained sequences were analysed using BLAST (Basic Local Alignment Search Tool, www.ncbi.nlm.nih.gov/BLAST/). All cases were tested for expression of a zinc-finger gene suppressor of zeste 12 homolog (Drosophila) (*SUZ12*) to assess RNA quality.

Long Distance PCR

PCR experiments on genomic DNA were performed using ~100 ng DNA as template in 25 µl PCR reactions using TaKaRa LA Taq following the manufacturer's recommendations for LD-PCR: 30 cycles of 98°C for 10 sec (denaturation) followed by 68°C for 15 min (annealing and extension; Takara Bio Inc). PCR products were purified using GeneJET PCR purification kit (Fermentas GmbH, St. Leon-Rot, Germany) and sent for Sanger sequencing (GATC Biotech).

Table 1. Patient characteristics.

Patient number	Sex/age at diagnosis	Histological diagnosis	Location of primary tumour	Tumour material analysed
1	F/63	Mesenchymal CS	Left iliacus muscle (soft tissue tumour)	Primary tumour
2	F/38	Mesenchymal CS	Pelvic bone (bone tumour)	Metastasis
3	F/12	Mesenchymal CS	Vertebra (bone tumour)	Primary tumour
4	F/39	Mesenchymal CS	Vertebra and right thigh (bone and soft tissue tumour)	Metastasis

F = female.
doi:10.1371/journal.pone.0049705.t001

Table 2. List of primers.

Primers	Sequence (5′ to 3′)	Tested fusion
HEY1_F1	CGAGGTGGAGAAGGAGAGTG	*HEY1-NCOA2*
NCOA2_E13-R3	AGTTGGGCTTTGCAATGTGA	*HEY1-NCOA2*
CDX1-214F	CCGCAGTACCCCGACTTCTCCAG	*CDX1-IRF2BP2*
CDX1-369F	ATTCGGGCCCCCTCCAGACTTTA	*CDX1-IRF2BP2*
CDX1-659R	GTTCAGTGAGCCCCAGATTGGCAG	*IRF2BP2- CDX1*
CDX1-771R	TGATGTCGTGGGCCATCGGC	*IRF2BP2- CDX1*
CDX1-26970R	GTCTCAGGCTCCCCTTCGTGAGTGTGTC	*IRF2BP2- CDX1*
IRF2BP2-895F	CAAGAGCCGCGGGTCTGGAGA	*IRF2BP2- CDX1*
IRF2BP2-926F	GTCAACAGGCCCAAGACCGTGC	*IRF2BP2- CDX1*
IRF2BP2-1172R	CTTGAGCCCCTCTGTGGATGTGGA	*CDX1-IRF2BP2*
IRF2BP2-1248R	GTGTGGTCCGGTTGGAATGAGGTG	*CDX1-IRF2BP2*

doi:10.1371/journal.pone.0049705.t002

Results

The cytogenetic analysis of the only tumour (patient 1) from which we got a fresh sample revealed a balanced t(1;5)(q42;q32) as the sole abnormality in all cells analysed (Figure 2A). Analysis of peripheral blood leukocytes displayed a normal female karyotype ruling out the possibility of a constitutional aberration.

All cases (i.e., the tumour carrying a t(1;5) as well as the three archival tumours) were tested for the *HEY1-NCOA2* fusion which was recently identified in 10 of 15 investigated MCs [10]. A PCR product of approximately 300 bp was amplified from tumours 2, 3, and 4 but not from tumour 1. Subsequent sequencing analysis confirmed the fusion between *HEY1* exon 4 and *NCOA2* exon 13 in cases 2–4, identical to the one previously described [10] (Figures 1B and 1C). As no PCR product was amplified in tumour 1, i.e., the one showing the t(1;5), we assumed that a new fusion gene was generated by the 1;5-rearrangement. To better characterize the breakpoint on the rearranged chromosome 5, a series of selected BAC clones mapping to the involved bands were hybridized to metaphase plates. Clone CTC-802J2 mapping on 5q32 and covering four genes gave three signals on metaphase chromosomes. The breakpoint position was further narrowed down using fosmid clones with clone G248P81640F4 giving a split signal, mapping the breakpoint to a genomic area between the 3′end of the platelet derived growth factor receptor β gene (*PDGFRβ*) and the large intron 1 of *CDX1*.

Since the breakpoint region as delimited by FISH was large, probably involving one of two genes, we decided to investigate the translocation in detail using a whole transcriptome sequencing approach focusing on potential fusion transcripts between chromosomes 1 and 5. The deFuse algorithm [15], designed for fusion gene discovery in paired-end RNA sequence data sets, gave us a list of 92 putative fusions in the tumour transcriptome that we reduced to 85 after removing isoforms of the same fusions (Table S2). An *IRF2BP2-CDX1* transcript involving two coding regions yielded the highest split count value (number of split reads supporting the fusion) of all the predicted fusions and was predicted to be in-frame. We therefore chose to focus on this putative fusion. cDNA PCR experiments with specific primers were run to validate the *IRF2BP2-CDX1* fusion. Two distinct bands were identified using the primer IRF2BP2-895F, located in exon 1, in combination with CDX1-771R, located in exon 3 (Table 2 and Figure 2C). The primer combination IRF2BP2-926F and CDX1-659R yielded smaller but similar bands. Cloning of the amplified PCR products was performed and sequencing was carried out from six individual bacterial clones. *IRF2BP2* exists as two different isoforms where isoform B lacks 48 bp of exon 1 sequence representing 16 amino acids [17]. Sequencing analysis of the PCR products confirmed the presence of both isoforms fused to exon 2 of *CDX1*, i.e., an *IRF2BP2-CDX1* fusion transcript was confirmed in the tumour RNA. Both fusion transcripts were found to be "in frame" and are predicted to encode proteins of 466 and 450 amino acids, respectively, before being terminated by a stop codon. The reciprocal fusion between *CDX1-IRF2BP2* did not yield any products by cDNA PCR (primer combinations CDX1-214F+IRF2BP2-1248R and CDX1-369F+IRF2BP2-1172R, Table 2). The presence of the *IRF2BP2-CDX1* fusion gene was tested for in specimens from tumour 2–4 using primer combinations IRF2BP2-926F and CDX1-771R (Table 2). None of the specimens showed such fusion.

Next, we wanted to identify the genomic breakpoints of the *IRF2BP2-CDX1* fusion. Using primers IRF2BP2-926F and CDX1-26970R (Table 2) we managed to amplify a product of about 800 bp which by sequencing was shown to contain the breakpoint, i.e., sequencing analysis confirmed the predicted positions of the genomic breakpoints. On chromosome 1 the breakpoint was in intron 1 of *IRF2BP2* (chr1:234743757 bp), whereas on chromosome 5 it was in the large intron 1 of *CDX1* (chr5:149551799 bp).

To investigate the involvement of the *IRF2BP2* gene also by FISH, we hybridized BAC clones overlapping *IRF2BP2* to metaphases obtained from the cultured cells (Table S1). Signals were detected on the normal chromosome 1 and the derivative chromosome 1. However, no signal was seen on chromosome 5 as would be expected if the *IRF2BP2-CDX1* fusion had resulted from a simple balanced translocation. These findings thus indicate that a more complex rearrangement had taken place, possibly including inversions at the breakpoint.

Discussion

The cytogenetic knowledge on mesenchymal chondrosarcomas is limited to ten cases [9]. We report here a solitary t(1;5)(q42;q32) in a case of MC. The translocation led to recombination of the *IRF2BP2* and *CDX1* genes.

This is the first time an *IRF2BP2-CDX1* fusion has been detected in human neoplasia. *CDX1* belongs to the homeobox gene family [18]. These genes share a homebox domain that encodes a DNA binding protein functioning as a transcription

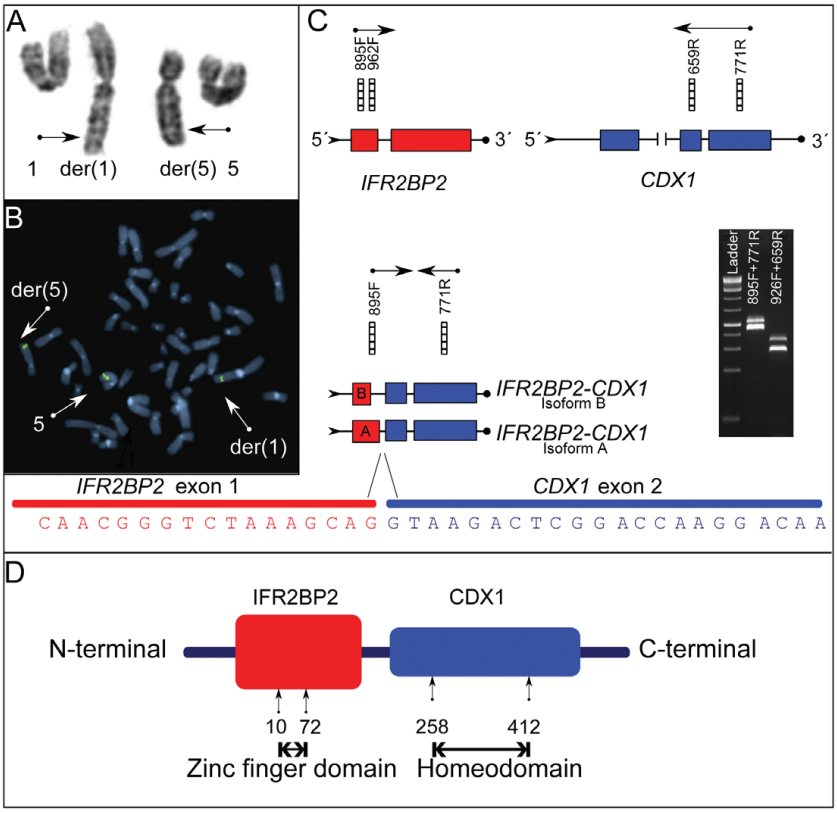


Figure 2. Cytogenetic and molecular details of the *IRF2BP2- CDX1* fusion gene. (A) Partial karyotype showing the aberrant chromosomes 1 and 5. Arrows point to the breakpoint positions. (B) Inverted DAPI metaphase harbouring the t(1;5). Upon hybridization with probe CTC-802J2 mapping to 5q32, three fluorescent signals were detected; (on the normal chromosome 5, the derivative chromosome 5, and the derivative chromosome 1), indicating a breakpoint within the genomic area covered by this BAC. (C) In the upper panel, the structure of the wild type *IRF2BP2* and *CDX1* genes is shown in grey and black, respectively. Bars indicate positions of primers yielding products by cDNA PCR. For detailed primer information, see Table 2. In the lower panel, the two identified fusion gene transcripts are illustrated. By sequencing the fusion was found to consist of *IRF2BP2* exon 1 (isoform A or B) fused to exon 2 of *CDX1*. The base sequence shown originates from isoform A. A gel blot demonstrating the two PCR products is shown in the right panel. The primer combinations used are specified. (D) Illustration of hypothetical fusion protein. The N terminal part of the protein originates from IRF2BP2 and harbour a zinc finger motif which may bind DNA [17]. The C terminal part contains a homeodomain which may also interact with the DNA double helix (http://www.ncbi.nlm.nih.gov/gene/1044, accessed August 2012).
doi:10.1371/journal.pone.0049705.g002

factor [19]. In particular, CDX1 is an upstream regulator of Hox-gene expression [20] that has been implicated in malignancies such as leukaemias [21,22]. In adults, CDX1 expression is restricted to intestinal epithelium [23–25] and aberrant expression has been linked to intestinal cancer [24,26–28]. No fusion gene involving *CDX1* has so far been described as opposed to another member of the Cdx family, *CDX2*. *CDX2* is overexpressed in both lymphoid and myeloid leukaemias [29–31] and a fusion gene resulting from a balanced t(12;13) leading to an *ETV6-CDX2* fusion was detected in a patient with acute myeloid leukaemia [32].

The first exon of *IRF2BP2* forms the 5end of the *IRF2BP2-CDX1* fusion. *IRF2BP2* normally exists in two isoforms resulting from alternative splicing of the gene [17]. Both variants contain a Zinc finger motif at their N-terminus possibly binding DNA [17]. Although no direct link to cancer has been described for this gene, *IRF2BP2* interacts with partners that are involved in cancer as for example the tumour suppressor gene *TP53* [33] and the oncogene *IRF2* [17]. *IRF2BP2* also acts as a co-repressor of *IRF2*, inhibiting the expression of interferon-responsive genes. Recently also *NFAT1*, which encodes a transcription factor involved in the cell cycle, differentiation, and apoptosis, was shown to be repressed by *IRF2BP2* [34]. According to BioGPS [35], *IRF2BP2* is expressed in a variety of human tissues [36].

Two PCR products were obtained by cDNA PCR investigations for the *IRF2BP2-CDX1* fusion. The difference between the two products was by sequencing shown to be caused by the alternative splice variants of *IRF2BP2*. Both sequences were shown to be in frame, with the largest transcript predicted to encode a 466 amino acid protein and the smaller encoding 450 amino acids. The

biological implications of the predicted fusion protein IRF2BP2-CDX1 can only be speculated upon, but as both fusion partners are involved in transcriptional regulation, a protein disturbing DNA transcription is likely. The *IRF2BP2-CDX1* fusion is thus suggested to take part in MC tumorigenesis and/or progression.

MCs are rare, and at our institution only five patients (four included in this study) received this diagnosis during the last 25 years. We identified the *HEY1-NCOA2* in three of these tumours (patients 2, 3, and 4), confirming that this fusion gene is common in MC. In a majority of the well-known translocation-related sarcomas such as myxoid liposarcoma and low-grade fibromyxoid sarcoma, more than one defining fusion variant has been detected. Often one fusion variant is more common than the others [2,3]. *IRF2BP2-CDX1* could be such an additional fusion variant identified in a subset of MCs, but only analysis of larger series of tumours can determine the prevalence of the *IRF2BP2-CDX1*.

Of possible interest is the fact that the three *HEY1-NCOA2*-positive MCs all had tumour manifestations detected in bone, whereas the MC showing the t(1;5) and *IRF2BP2-CDX1* fusion originated from soft tissue. Although most common in bone, one-fifth to one-third of MCs do arise in soft tissue [4]. The tissue of manifestation was not reported in the MCs where the *HEY1-NCOA2* was first described [10]. Given the rarity of these tumours, only future surveys of larger groups of patients can clarify if there is a correlation between the tissue the tumour affects and the type of fusion gene present. This study demonstrates the feasibility and indeed advantage of using karyotyping and molecular cytogenetic methods together with transcriptome sequencing to identify fusion genes caused by chromosomal rearrangements. Traditionally, chromosome walking using BACs or equivalent probes has been used to narrow down the breakpoint regions followed by PCR based analyses to amplify the genes involved in the breakpoints. Submicroscopic rearrangements in the breakpoint area can cause considerable confusion, however, and prevent amplification of fusion genes. Using whole-transcriptome sequencing without prior genetic knowledge of the tumour investigated can also be challenging as validation of numerous predicted fusion gene transcripts is necessary. To know which chromosomes take part in the rearrangement therefore helps considerably when looking for novel putative cancer-specific fusion genes.

Supporting Information

Table S1 BAC probes used for FISH experiments in case 1.
(XLS)

Table S2 List of fusions suggested by the deFuse algorithm.
(XLS)

Acknowledgments

The authors would like to thank Kirsten Sundby-Hall for help handling register data. Researchers interested in more data are encouraged to contact the authors directly.

Author Contributions

Conceived and designed the experiments: FM SH RAL RIS. Performed the experiments: KBN IP JT LH LG MG TN. Analyzed the data: KBN IP JT LG TN MG RIS LH. Wrote the paper: KBN IP JT LG TN RIS FM SH BB AF. Provided essential diagnostic information: BB AF. Revised the manuscript critically: FM RAL SH.

References

1. Fletcher CDM, Unni KK, Mertens F, editors (2002) World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Soft Tissue and Bone. Lyon: IARCPress.
2. Mertens F, Antonescu CR, Hohenberger P, Ladanyi M, Modena P, et al. (2009) Translocation-related sarcomas. Semin Oncol 36: 312–323.
3. Mitelman F, Johansson B, Mertens F (2007) The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 7: 233–245.
4. Nakashima Y, Park JH, Sugano O (2002) Mesenchymal chondrosarcoma. In: Fletcher C, Unni K, Mertens F, editors. WHO classification of solide and soft tissue tumors. Lyon: IARCPress.
5. Nakashima Y, Unni KK, Shives TC, Swee RG, Dahlin DC (1986) Mesenchymal chondrosarcoma of bone and soft tissue. A review of 111 cases. Cancer 57: 2444–2453.
6. Riedel RF, Larrier N, Dodd L, Kirsch D, Martinez S, et al. (2009) The clinical management of chondrosarcoma. Curr Treat Options Oncol 10: 94–106.
7. Dantonello TM, Int-Veen C, Leuschner I, Schuck A, Furtwaengler R, et al. (2008) Mesenchymal chondrosarcoma of soft tissues and bone in children, adolescents, and young adults: experiences of the CWS and COSS study groups. Cancer 112: 2424–2431.
8. Cesari M, Bertoni F, Bacchini P, Mercuri M, Palmerini E, et al. (2007) Mesenchymal chondrosarcoma. An analysis of patients treated at a single institution. Tumori 93: 423–427.
9. Mitelman F, Johansson B, Mertens F, (Eds) (2012) Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. Available: http://cgap.nci.nih.gov/Chromosomes/Mitelman. Accessed 2012 May.
10. Wang L, Motoi T, Khanin R, Olshen A, Mertens F, et al. (2012) Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data. Genes Chromosomes Cancer 51: 127–139.
11. Abrey LE, Yahalom J, DeAngelis LM (2000) Treatment for primary CNS lymphoma: the next step. J Clin Oncol 18: 3144–3150.
12. Mandahl N (2001) Methods in solid tumour cytogenetics. In: Rooney DE, editor. Human Cytogenetics: Malignancy and acquired abnormalities. 3 ed. New York: Oxford University Press.
13. Shaffer LG, Slovak ML, Campbell LJ, editors (2009) ISCN 2009: An International System for Human Cytogenetic Nomenclature (2009). Basel: S. Karger.
14. Nyquist KB, Thorsen J, Zeller B, Haaland A, Troen G, et al. (2011) Identification of the TAF15-ZNF384 fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12). Cancer Genet 204: 147–152.
15. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, et al. (2011) deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol 7: e1001138.
16. Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. Genomics 98: 137–144.
17. Childs KS, Goodbourn S (2003) Identification of novel co-repressor molecules for Interferon Regulatory Factor-2. Nucleic Acids Res 31: 3016–3026.
18. Holland PW, Booth HA, Bruford EA (2007) Classification and nomenclature of all human homeobox genes. BMC Biol 5: 47.
19. Treisman J, Harris E, Wilson D, Desplan C (1992) The homeodomain: A new face for the helix-turn-helix? BioEssays 14: 145–150.
20. Lohnes D (2003) The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. BioEssays 25: 971–980.
21. Shah N, Sukumar S (2010) The Hox genes and their roles in oncogenesis. Nat Rev Cancer 10: 361–371.
22. Grier DG, Thompson A, Kwasniewska A, McGonigle GJ, Halliday HL, et al. (2005) The pathophysiology of HOX genes and their role in cancer. J Pathol 205: 154–171.
23. Bonner CA, Loftus SK, Wasmuth JJ (1995) Isolation, characterization, and precise physical localization of human CDX1, a caudal-type homeobox gene. Genomics 28: 206–211.
24. Silberg DG, Furth EE, Taylor JK, Schuck T, Chiou T, et al. (1997) CDX1 protein expression in normal, metaplastic, and neoplastic human alimentary tract epithelium. Gastroenterology 113: 478–486.
25. Mallo GV, Rechreche H, Frigerio JM, Rocha D, Zweibaum A, et al. (1997) Molecular cloning, sequencing and expression of the mRNA encoding human Cdx1 and Cdx2 homeobox. Down-regulation of Cdx1 and Cdx2 mRNA expression during colorectal carcinogenesis. Int J Cancer 74: 35–44.
26. Kang JM, Lee BH, Kim N, Lee HS, Lee HE, et al. (2011) CDX1 and CDX2 expression in intestinal metaplasia, dysplasia and gastric cancer. J Korean Med Sci 26: 647–653.
27. Chan CW, Wong NA, Liu Y, Bicknell D, Turley H, et al. (2009) Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. Proc Natl Acad Sci U S A 106: 1936–1941.
28. Wong NA, Britton MP, Choi GS, Stanton TK, Bicknell DC, et al. (2004) Loss of CDX1 expression in colorectal carcinoma: promoter methylation, mutation, and loss of heterozygosity analyses of 37 cell lines. Proc Natl Acad Sci U S A 101: 574–579.

29. Riedt T, Ebinger M, Salih HR, Tomiuk J, Handgretinger R, et al. (2009) Aberrant expression of the homeobox gene CDX2 in pediatric acute lymphoblastic leukemia. Blood 113: 4049–4051.
30. Scholl C, Bansal D, Dohner K, Eiwen K, Huntly BJ, et al. (2007) The homeobox gene CDX2 is aberrantly expressed in most cases of acute myeloid leukemia and promotes leukemogenesis. J Clin Invest 117: 1037–1048.
31. Frohling S, Scholl C, Bansal D, Huntly BJ (2007) HOX gene regulation in acute myeloid leukemia: CDX marks the spot? Cell Cycle 6: 2241–2245.
32. Chase A, Reiter A, Burci L, Cazzaniga G, Biondi A, et al. (1999) Fusion of ETV6 to the caudal-related homeobox gene CDX2 in acute myeloid leukemia with the t(12;13)(p13;q12). Blood 93: 1025–1031.
33. Koeppel M, van Heeringen SJ, Smeenk L, Navis AC, Janssen-Megens EM, et al. (2009) The novel p53 target gene IRF2BP2 participates in cell survival during the p53 stress response. Nucleic Acids Res 37: 322–335.
34. Carneiro FR, Ramalho-Oliveira R, Mognol GP, Viola JP (2011) Interferon regulatory factor 2 binding protein 2 is a new NFAT1 partner and represses its transcriptional activity. Mol Cell Biol 31: 2889–2901.
35. Su A, Wu C, Macleod I (2012) BioGPS. Available: http://biogps.org/#goto=genereport&id=359948. Accessed 2012 May.
36. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062–6067.

Paper IV

Francesca Micci, Jim Thorsen, Ioannis Panagopoulos, Kaja B. Nyquist, Bernward Zeller, Anne Tierens, Sverre Heim:

High-throughput sequencing identifies an *NFIA/CBFA2T3* fusion gene in acute erythroid leukemia with t(1;16)(p31;q24).

Letter to Leukemia (2013) 27(4):980–982.

Open

High-throughput sequencing identifies an *NFIA/CBFA2T3* fusion gene in acute erythroid leukemia with t(1;16)(p31;q24)

Leukemia (2013) **27,** 980–982; doi:10.1038/leu.2012.266

In a previous publication of ours[1] we showed the involvement of the myeloid translocation gene-related protein 2 gene (*CBFA2T3*) in a case of acute erythroid leukemia with the translocation t(1;16)(p31;q24). Because of lack of material available for analysis, we could not with certainty determine the leukemogenic mechanism, whether it be generation of a fusion gene with *CBFA2T3* as one of the partners or loss of tumor suppressor activity, in which case genes *KANK1*and *L1TD1*, both homozygously lost, might be of the essence.

Some months after our article was published we got hold of 2 ml coagulated blood with 3% of abnormal cells from the same patient. As sequencing technology has also developed very fast lately, we decided to extract RNA from the sample and try to use it for high-throughput sequence analysis. The RNA was extracted and its quality checked by the Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA). A total of 3 μg of RNA was sent for high-throughput pair-end RNA-sequencing to the Norwegian Sequencing Center at the Ullevål Hospital (http://www.sequencing.uio.no/). The Illumina software pipeline was used to process image data into raw sequencing data and only sequence reads marked as 'passed filtering' were used in the downstream data analysis. A total of 107 million reads were obtained. The FASTQC software was used for quality control of the raw sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We used the fusion discovery software FusionMap (release date 16-04-2012)[2] and the pre-built Human B37 and RefGene from the FusionMap website (http://www.omicsoft.com/fusionmap/). A list of over 500 possible fusion genes was obtained. A specific fusion involving the *CBFA2T3* gene, which maps to chromosome band 16q24, and the nuclear factor I/A (*NFIA*) gene, which maps to chromosome band 1p31, was identified as number 10 in the list (seed count-rank 48; Supplementary Table 1). The involvement of the *NFIA* gene fits well with the fluorescence *in situ* hybridization (FISH) data on chromosome 1 previously obtained and published.[1] The presence of the *NFIA/CBFA2T3* fusion was verified by PCR using the NFIA-956 F and CBFA2T3–622 R primer combination (Supplementary Table 2). A specific PCR product of about 500 bp was identified and directly sequenced (Figure 1a). The specific fusion occurs between exon 6 of the *NFIA* gene (accession number NM_001134673.3) and exon 3 of the *CBFA2T3* gene (accession number NM_005187.5; Figure 1b).

The *NFIA/CBFA2T3* fusion gave an open reading frame and is expected to lead to a chimeric protein containing 208 amino-acid residues from NFIA (according to NP_001128145.1) and 603 residues from CBFA2T3 (according to NP_005178.4). The predicted fusion protein should thus consist of 811 amino acids (Figure 1c).

The *NFIA* gene encodes a member of the NFI family of transcription factors (http://genome.ucsc.edu). Interestingly, it has been found that *NFIA* exhibits a marked lineage-specific expression pattern in normal human hematopoiesis; it is upregulated in the erythroid lineage but fully suppressed in granulocytopoiesis.[3] It has been shown that in early hematopoiesis, the NFIA expression level acts as a factor channeling hematopoietic progenitor cells into either the erythroid or granulocytopoietic lineage.[3] The NFI proteins have a DNA-binding and dimerization domain in their N-terminal half, which contains four cysteine residues, and a transactivation and repression domain in their C-terminal half.[4] The *NFIA* gene was found involved in an *NFIA/EHF* chimeric fusion in one breast cancer cell line out of 24 breast tumors analyzed (nine cell lines and 15 primary tumors).[5] However, its role as either a passenger event or a direct, albeit infrequent, contributor to breast cancer development remains uncertain.

CBFA2T3 encodes an ETO myeloid translocation gene family protein, which interacts with DNA-bound transcription factors and recruits a variety of corepressors to facilitate transcriptional repression.[6–8] The t(16;21)(q24;q22) translocation is one of the less common karyotypic abnormalities specifically associated with acute myeloid leukemia (AML). The translocation produces a chimeric gene made up of the 5′-region of the runt-related transcription factor 1 (*RUNX1*) gene fused to the 3′-region of *CBFA2T3* (Figure 1d). In AMLs with either t(8;21) or t(16;21), the transcription factor RUNX1 is juxtaposed to one of the zinc finger nuclear proteins CBFA2T1 and CBFA2T3, respectively, resulting in transcriptional repression of *RUNX1* target genes.[6] Lately, its involvement as a partner in fusion genes was underlined by the identification of a *IGH/CBFA2T3* fusion in a case of Burkitt lymphoma and a diffuse large B-cell lymphoma.[9] This gene is also a putative breast tumor suppressor.[10,11] Interestingly, *CBFA2T3* is downregulated during erythroid differentiation, and it has been

Accepted article preview online 11 September 2012; advance online publication, 2 October 2012

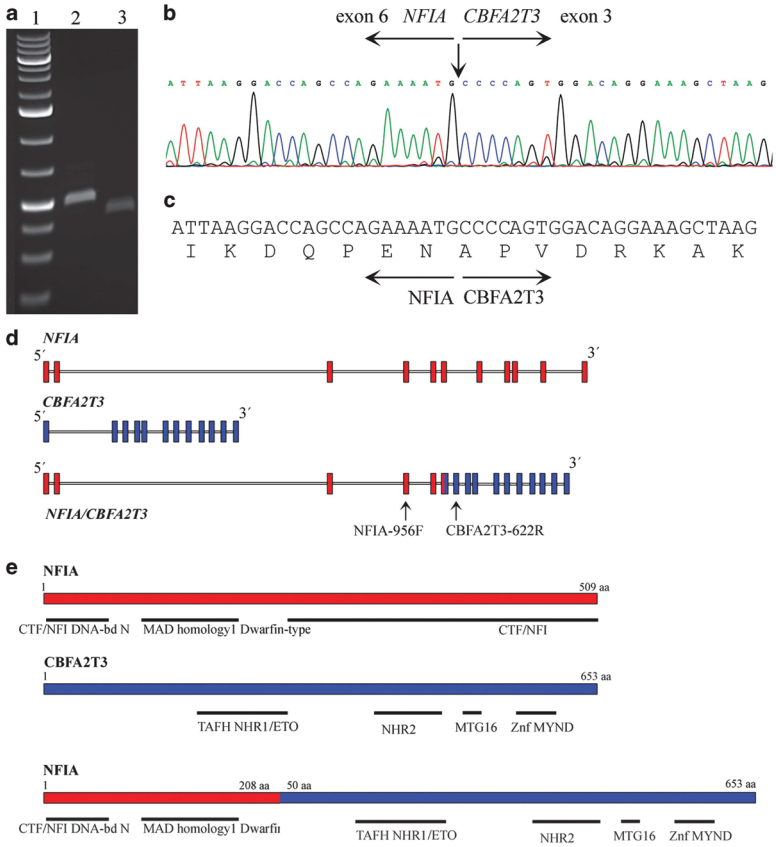


Figure 1. Detection of the *NFIA/CBFA2T3* fusion. (**a**) Gel picture showing the amplified fragment. Lane 1: ladder, lane 2: PCR product obtained with primer combination NFIA-924 F and CBFA2T3–622 R, lane 3: product of the NESTED-PCR obtained with primer combination NFIA-956 F and CBFA2T3–598 R. (**b**) Partial chromatogram showing the junction of the *NFIA* and *CBFA2T3* genes. (**c**) Deduced amino acid sequence of the fusion transcript. (**d**) Schematic overview of the breakpoint region of the *NFIA* and *CBFA2T3* genes. The exons are not in scale. Arrows point to primer positions. (**e**) Schematic overview of the position of the different domains of the NFIA and CBFA2T3 proteins and the NFIA/CBFA2T3 chimeric protein, according to ensembl (http://www.ensembl.org/index.html).

suggested to have a repressive role in early, as well as late human erythroid differentiation.[12] Hildebrand *et al.*[6] demonstrated that the nuclear protein ETO (eight-twenty-one, a family to which also CBFA2T1, CBFA2T2 and CBFA2T3 belong) does not show reduced repressor activity even if it lacks the first 236 amino acids. As in the present fusion the altered CBFA2T3 protein lacks only the first 50 amino acids, we assume that its repressor activity is still retained (Figure 1e). More specifically, we hypothesize a pathogenetic parallel between AML showing a t(8;21) or t(16;21) and the present erythroleukemia with the 1;16-translocation with transcriptional repression of the *NFIA* target genes in the present case.

As the karyotype was described as 46,XY,der(1)t(1;1)(p31;q21),-del(1)(p11p31),der(16)t(1;16)(p31;q24), that is, presented additional rearrangement besides the 1;16-translocation, we decided to screen the list of possible fusion genes in search of genes located in karyotypic breakpoints to see if those were involved in fusions as well. We identified four possible fusions (seed count-rank >12) where one of the genes mapped to a breakpoint position on chromosome 1. An analysis of the hypothetical fusions using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) showed in one of the two genes high-sequence identity with several genes and/or numerous repetitive sequences (for example, SINE). Hence, the reality of the putative fusions was seriously called into question and no further investigations were undertaken.

In addition to the present case, two more cases of erythroleukemia showing a t(1;16)(p31;q24) in their karyotype[13,14] can be found in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer.[15] All three patients (including ours) were very young children, and clinical outcome was poor. We assume that a *NFIA/CBFA2T3* fusion existed also in these leukemias, but no evidence is at hand to corroborate or falsify this assumption.

In summary, we describe the first fusion gene identified in acute erythroleukemia. Knowledge of its specific functions in the neoplastic context is still incomplete, but pathogenetic similarities with other leukemic fusion genes are readily discernible. As for other leukemias characterized genetically by fusion genes, one may assume that the detailed pathogenetic knowledge now emerging may eventually form a starting point from which therapeutic attempts may begin.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was approved by the Regional Ethics Committee (REK number: S-07474a) and the institutional review board. The study was supported by grants from the Norwegian Cancer Society and the South-Eastern Norway Regional Health Authority. The sequencing service was provided by the Norwegian High-Throughput Sequencing Center, a national technology platform supported by the 'Functional Genomics' and 'Infrastructure' programs of the Research Council of Norway and the South-Eastern Regional Health Authorities.

F Micci[1,2], J Thorsen[1,2], I Panagopoulos[1,2], KB Nyquist[1,2], B Zeller[3], A Tierens[4] and S Heim[1,2,5]
[1]*Section for Cancer Cytogenetics, Institute for Medical Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway;*
[2]*Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway;*
[3]*Department of Pediatrics, Rikshospitalet, Oslo University Hospital, Oslo, Norway;*
[4]*Department of Pathology, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway and*
[5]*Faculty of Medicine, University of Oslo, Oslo, Norway*
E-mail: francesca.micci@labmed.uio.no

REFERENCES

1 Micci F, Thorsen J, Haugom L, Zeller B, Tierens A, Heim S. Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. *Leukemia* 2011; **25**: 1510–1512.
2 Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* 2011; **27**: 1922–1928.
3 Starnes LM, Sorrentino A, Pelosi E, Ballarino M, Morsilli O, Biffoni M *et al.* NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression. *Blood* 2009; **114**: 1753–1763.
4 Bernard F, Gelsi-Boyer V, Murati A, Giraudier S, Trouplin V, Adelaide J *et al.* Alterations of NFIA in chronic malignant myeloid diseases. *Leukemia* 2009; **23**: 583–585.
5 Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009; **462**: 1005–1010.
6 Hildebrand D, Tiefenbach J, Heinzel T, Grez M, Maurer AB. Multiple regions of ETO cooperate in transcriptional repression. *J Biol Chem* 2001; **276**: 9889–9895.
7 Kumar R, Cheney KM, Neilsen PM, Schulz RB, Callen DF. CBFA2T3-ZNF651, like CBFA2T3-ZNF652, functions as a transcriptional corepressor complex. *FEBS Lett* 2010; **584**: 859–864.
8 Kumar R, Selth LA, Schulz RB, Tay BS, Neilsen PM, Callen DF. Genome-wide mapping of ZNF652 promoter binding sites in breast cancer cells. *J Cell Biochem* 2011; **112**: 2742–2747.
9 Salaverria I, Akasaka T, Gesk S, Szczepanowski M, Burkhardt B, Harder L *et al.* The CBFA2T3/ACSF3 locus is recurrently involved in IGH chromosomal translocation t(14;16)(q32;q24) in pediatric B-cell lymphoma with germinal center phenotype. *Genes Chromosomes Cancer* 2012; **51**: 338–343.
10 Kochetkova M, McKenzie OL, Bais AJ, Martin JM, Secker GA, Seshadri R *et al.* CBFA2T3 (MTG16) is a putative breast tumor suppressor gene from the breast cancer loss of heterozygosity region at 16q24.3. *Cancer Res* 2002; **62**: 4599–4604.
11 Kumar R, Manning J, Spendlove HE, Kremmidiotis G, McKirdy R, Lee J *et al.* ZNF652, a novel zinc finger protein, interacts with the putative breast tumor suppressor CBFA2T3 to repress transcription. *Mol Cancer Res* 2006; **4**: 655–665.
12 Kiefer CM, Lee J, Hou C, Dale RK, Lee YT, Meier ER *et al.* Distinct Ldb1/NLI complexes orchestrate gamma-globin repression and reactivation through ETO2 in human adult erythroid cells. *Blood* 2011; **118**: 6200–6208.
13 Koller U, Haas OA, Ludwig WD, Bartram CR, Harbott J, Panzer-Grumayer R *et al.* Phenotypic and genotypic heterogeneity in infant acute leukemia. II. Acute nonlymphoblastic leukemia. *Leukemia* 1989; **3**: 708–714.
14 Castaneda VL, Parmley RT, Saldivar VA, Cheah MS. Childhood undifferentiated leukemia with early erythroid markers and c-myb duplication. *Leukemia* 1991; **5**: 142–149.
15 Mitelman F, Johansson B, Mertens F. Mitelman database of chromosome aberrations and gene fusions in cancer. http://cgap nci nih gov/Chromosomes/Mitelman 2012. Available from: URL http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)